Image Classification of High Variant Objects in Fast Industrial Applications

Vom Fachbereich Informatik
der Technischen Universität Darmstadt
genehmigte Dissertation

DISSERTATION

zur Erlangung des akademischen Grades eines
Doktor-Ingenieurs (Dr.-Ing.)
von

M.A. Dirk Siegmund

Potsdam, Deutschland

Referenten der Arbeit: Prof. Dr. Arjan Kuijper
Technische Universität Darmstadt
Prof. Dr. techn. Dr.-Ing. eh. Dieter W. Fellner
Technische Universität Darmstadt
Prof. Dr. Fabrizzio Soares
Universidade Federal de Goiás

Tag der Einreichung: 20/09/2023

Tag der mündlichen Prüfung: 13/11/2023

Darmstädter Dissertation
D 17
Darmstadt 2023

Siegmund, Dirk: Image Classification of High Variant Objects in Fast Industrial Applications
Darmstadt, Technische Universität Darmstadt,

Jahr der Veröffentlichung der Dissertation auf TUprints: 2024
Tag der mündlichen Prüfung: 13.11.2023

Abstract

Recent advances in machine learning and image processing have expanded the applications of computer vision in many industries. In industrial applications, image classification is a crucial task since high variant objects present difficult problems because of their variety and constant change in attributes. Computer vision algorithms can function effectively in complex environments, working alongside human operators to enhance efficiency and data accuracy. However, there are still many industries facing difficulties with automation that have not yet been properly solved and put into practice. They have the need for more accurate, convenient, and faster methods. These solutions drove my interest in combining multiple learning strategies as well as sensors and image formats to enable the use of computer vision for these applications. The motivation for this work is to answer a number of research questions that aim to mitigate current problems in hinder their practical application. This work therefore aims to present solutions that contribute to enabling these solutions. I demonstrate why standard methods cannot simply be applied to an existing problem. Each method must be customized to the specific application scenario in order to obtain a working solution.

One example is face recognition where the classification performance is crucial for the system's ability to correctly identify individuals. Additional features would allow higher accuracy, robustness, safety, and make presentation attacks more difficult. The detection of attempted attacks is critical for the acceptance of such systems and significantly impacts the applicability of biometrics. Another application is tailgating detection at automated entrance gates. Especially in high security environments it is important to prevent that authorized persons can take an unauthorized person into the secured area. There is a plethora of technology that seem potentially suitable but there are several practical factors to consider that increase or decrease applicability depending which method is used. The third application covered in this thesis is the classification of textiles when they are not spread out. Finding certain properties on them is complex, as these properties might be inside a fold, or differ in appearance because of shadows and position.

The first part of this work provides in-depth analysis of the three individual applications, including background information that is needed to understand the research topic and its proposed solutions. It includes the state of the art in the area for all researched applications. In the second part of this work, methods are presented to facilitate or enable the industrial applicability of the presented applications. New image databases are initially presented for all three application areas. In the case of biometrics, three methods that identify and improve specific performance parameters are shown. It will be shown how melanin face pigmentation (MFP) features can be extracted and used for classification in face recognition and PAD applications. In the entrance control application, the focus is on the sensor information with six methods being presented in detail. This includes the use of thermal images to detect humans based on their body heat, depth images in form of RGB-D images and 2D image series, as well as data of a floor mounted sensor-grid. For textile defect detection several methods and a novel classification procedure, in free-fall is presented.

In summary, this work examines computer vision applications for their practical industrial applicability and presents solutions to mitigate the identified problems. In contrast to previous work, the proposed approaches are (a) effective in improving classification performance (b) fast in execution and (c) easily integrated into existing processes and equipment.

Zusammenfassung

Jüngste Fortschritte im Bereich des maschinellen Lernens und der Bildverarbeitung haben die Anwendungsmöglichkeiten von Computer Vision in vielen Branchen erweitert. In industriellen Anwendungen ist die Bildklassifizierung eine wichtige Aufgabe, da variantenreiche Objekte aufgrund ihrer aufgrund ihrer Vielfalt und der ständigen Veränderung ihrer Eigenschaften schwierige Probleme darstellen. Bildverarbeitungsalgorithmen können in komplexen Umgebungen effektiv arbeiten und mit menschlichen Bedienern zusammenarbeiten, um Effizienz und Datengenauigkeit zu verbessern. Es gibt jedoch noch viele Branchen, die Lösungen benötigen, die noch nicht noch nicht richtig gelöst und in die Praxis umgesetzt wurden. Sie zeigen den Bedarf an genaueren, bequemeren und schnelleren Methoden. Diese Lösungen haben mein Interesse an der Kombination verschiedener Lernstrategien sowie Sensoren und Bildformaten geweckt um den Einsatz von Computer Vision für diese Anwendungen zu ermöglichen. Die Motivation für diese Arbeit ist die Suche nach Lösungen für eine Reihe von Forschungsfragen zu finden, die derzeit ihre praktische Anwendung behindern. Ziel dieser Arbeit ist es daher, Lösungen zu präsentieren, die zur Erreichung der Anwendbarkeit beitragen. Ich zeige, warum Standardmethoden nicht einfach auf ein bestehendes Problem angewendet werden können. Jede Methode muss an das jeweilige Anwendungsszenario angepasst werden, um eine funktionierende Lösung zu erhalten. Ein Beispiel ist die Klassifizierung von Textilien, wenn sie nicht ausgebreitet sind. Das Auffinden bestimmter Defekte ist sehr schwierig, da sich diese Eigenschaften in einer Falte befinden können oder aufgrund des sie umgebenden Schatten anders aussehen. Daher sind neue Ansätze notwendig um die Übertragbarkeit der Erkennungsalgorithmen auf 3D-Objekte zu verbessern. Eine weitere Anwendung ist die Gesichtserkennung, bei der die Klassifizierungsleistung entscheidend für die Fähigkeit des Systems ist, Personen korrekt zu identifizieren. Bislang ist nicht bekannt, ob die Melaminpigmentierung als zusätzliches biometrisches Merkmal verwendet werden kann. Wenn sie zusätzliche Merkmale hätte, würde sie eine höhere Genauigkeit, Robustheit und Sicherheit ermöglichen und Präsentationsangriffe erschweren. Die Erkennung von Angriffsversuchen ist für die Akzeptanz solcher Systeme von entscheidender Bedeutung und wirkt sich erheblich auf die Anwendbarkeit biometrischer Verfahren aus. Die dritte Anwendung, die in dieser Arbeit behandelt wird, ist die Erkennung von Zutrittsverletzungen an automatischen Eingangstoren. Gerade in Hochsicherheitsbereichen ist es wichtig zu verhindern, dass Unbefugte eine unbefugte Person in den gesicherten Bereich bringen können. Es gibt eine Fülle von Technologien, die potentiell geeignet erscheinen, aber es gibt mehrere praktische Faktoren zu berücksichtigen, die die Anwendbarkeit erhöhen oder verringern, je nachdem welcher Algorithmus verwendet wird. Der erste Teil dieser Arbeit enthält eine eingehende Analyse der drei einzelnen Anwendungen sowie Hintergrundinformationen, die zum Verständnis des Forschungsthemas dieser Arbeit und der vorgeschlagenen Lösungen erforderlich sind. Er umfasst den Stand der Technik in diesem Bereich für alle untersuchten Anwendungen.

Im zweiten Teil dieser Arbeit werden Methoden vorgestellt, die die Anwendbarkeit der vorgestellten Anwendungen für den industriellen Einsatz erhöhen bzw. ermöglichen. Für alle drei Anwendungsbereiche werden zunächst neue Bilddatenbanken vorgestellt. Im Falle der Biometrie werden drei Methoden gezeigt, die bestimmte Leistungsparameter identifizieren und verbessern. Es wird gezeigt, wie Melanin-Gesichtspigmentierungs-Merkmale (MFP) extrahiert und zur Klassifizierung in der Gesichtserkennung und in PAD-Anwendungen verwendet werden können. Bei der Eingangskontrollanwendung liegt der Schwerpunkt auf den Sensorinformationen, wobei sechs Methoden im Detail vorgestellt werden. Dazu gehören die Verwendung von Wärmebildern zur Erkennung

von Menschen anhand ihrer Körperwärme, Tiefenbilder in Form von RGB-D-Bildern und 2D-Bildserien sowie Daten eines am Boden montierten Sensorgitters.

Alle Ansätze werden dahingehend analysiert, ob sie die Anwendbarkeit zuverlässig erhöhen und die untersuchten Methoden somit industriell anwendbar sind. Zusammenfassend lässt sich sagen, dass in dieser Arbeit Bildverarbeitungsanwendungen auf ihre praktische Anwendbarkeit hin untersucht und Lösungen zur Entschärfung der festgestellten Probleme vorgestellt werden. Im Gegensatz zu früheren Arbeiten sind die vorgeschlagenen Ansätze (a) wirksam bei der Verbesserung der Klassifizierungsleistung, (b) schnell in der Ausführung und (c) leicht in bestehende Anlagen zu integrieren.

Contents

1. Introduction

1.1. Motivation

The world is currently experiencing a fourth industrial revolution. Following mechanization, electrification, and computerization, we are now witnessing a sustainable transformation in the industry through cyber-physical systems and smart factories. Traditional engineering sciences are merging with IT, resulting in the networking and integration of components, machines, and entire factories with new technologies like artificial intelligence (AI). As a result, entire process chains are now fully automated, revolutionizing manufacturing. While computer vision (CV) has been utilized in manufacturing for decades, recent advancements in machine learning and image processing have unlocked new possibilities. AI-assisted computer vision platforms are no longer limited to structured, repetitive tasks but can now function effectively in increasingly complex environments. These platforms work seamlessly alongside human operators, leading to enhanced efficiency, reduced errors, and improved data accuracy. However, it is important to note that not all AI platforms offer the same benefits. For optimal results, strategic integration of computer vision solutions into smart factories is crucial. By doing so, they can boost both digital and human performance on the production line. Historically, machine vision referred to a subset of manufacturing applications for computer vision technology, but with the advent of AI and advanced image processing, the scope of possibilities has expanded significantly.[Kla21]. Since the 1970s, systems of cameras, lights, reflectors, and software have helped manufacturers automate and optimize vision-based tasks. With robust applications in quality assurance, compliance, and inventory management, versions of machine vision have been crucial to manufacturing efficiently at scale. These systems, though, are limited to repetitive, structured scenarios and object surfaces that are easy to inspect. Their cost – between cameras, software, and tightly controlled lighting – can be restrictive [Chu17]. They were among the first systems to use early machine learning in manufacturing, but they require specialists to program and can be inflexible once deployed. Nevertheless, the first applications of image processing have been very successful and have become widespread. This includes detecting deviations from standards and automatically initiating corrective action or detecting quality defects in machine parts that are too subtle for the human eye. The main difference between what is known as machine vision systems and computer vision systems is flexibility. On top of processing structured information in controlled environments, computer vision systems have an enhanced ability to function in semi- and unstructured scenarios. And they're not limited to repetitive, automated processes. Industry 4.0 CV systems are better able to work with operators, responding to their actions and enabling new forms of digital-human interaction. Industry 4.0 describes the intelligent networking of machines and processes in industry with the help of information and communication technology.

Computer vision is a central sub-field of AI which has been developed very rapidly and applied to different manufacturing industries since such as food, pharmaceutical, automotive, aerospace, railway, semiconductor, electronic component, and other fields [LUNR15]. Vision-based industrial inspection was one driver of that research but other applications like biometrics also got much attention. The research on CV-based biometrics started in the late 1960s and early 1970s with the development of automated facial recognition systems. Since then, biometrics has expanded to include a wide range of techniques for recognizing and verifying human identity using different biometric traits. Common applications of computer vision are:

1. Object recognition and tracking: Identification and tracking of objects in real-time (as it is used in, e.g., autonomous vehicles or surveillance systems).
2. Quality control and inspection: Inspection of products during manufacturing to ensure they meet quality standards.
3. Biometrics: Recognition of biometric traits such as faces, fingerprints, or irises.

However, there are many more known applications than these three and other applications that may even be unknown. Computer vision is a key driver for Industry 4.0 because it enables machines to perceive, understand, and interact with the world around them. It is a crucial component of this revolution because it enables machines to process visual information in real-time, providing them with the ability to make decisions and take actions based on that information. The possibilities include, for example, flexible production, convertible factory, better logistics, and most importantly, the "use of data". Data on the production process and the condition of a product are brought together and evaluated. This is the basis for completely new business models and services.

With new techniques like "deep learning", machines can even recognize complex, coherent patterns automatically. One example is high variant objects which could not be classified sufficiently well just before the success of deep learning a few years ago. High variant objects are objects who are deformable and greatly differ in their appearance. Nevertheless, some industries demonstrate a need for CV solutions; however, many of them have not yet properly addressed how these solutions can be effectively applied in practice. One example is the classification of textiles when they are not spread out. Finding certain properties on them is very difficult, as these properties might be inside a fold, or they look different because of the shadow around them. This shows the need for more accurate, convenient, and large-scale approaches. These solutions drove my interest in combining multiple learning strategies as well as sensor and image formats in order to enable the use of computer vision in some applications. The motivation for this work is the challenge itself to find solutions to a number of research questions that still prevent the practical application of computer vision. This thesis therefore aims to present solutions that contribute to the achievement of applicability. In Chapter 1.2, general challenges that currently exist in the field of computer vision are first presented. In Section 1.3 I present the applications which are considered in this thesis and for which my contributions should bring improvements. At the end of each of the applications, I introduce the specific research questions on the respective topic that will be solved in this dissertation.

1.2. Research Challenges

In recent years, the field of computer vision has experienced remarkable advancements, revolutionizing various domains such as autonomous vehicles, medical imaging, robotics, and surveillance systems. Computer vision aims to enable machines to perceive and understand visual information, mimicking human vision capabilities. However, despite the significant progress made, numerous challenges persist, pushing researchers to delve deeper into this multidisciplinary field. This thesis delves into the intricacies of computer vision research challenges, exploring the complex problems that hinder the development of robust and reliable vision systems. As computer vision strives to achieve human-level perception, it must confront a myriad of obstacles, spanning from image acquisition and representation to high-level understanding and interpretation of visual data. Computer vision encompasses a number of sub-disciplines whose challenges are addressed by thousands of researchers worldwide. The most important research areas are listed in Figure 1.1.

One of the primary challenges faced by computer vision researchers lies in the domain of image and video understanding. While humans effortlessly interpret complex scenes and recognize objects, the same tasks prove daunting for machines. The inherent variability in images due to different lighting conditions, occlusions, and viewpoint changes poses significant hurdles in achieving accurate and consistent recognition. Extracting mean-

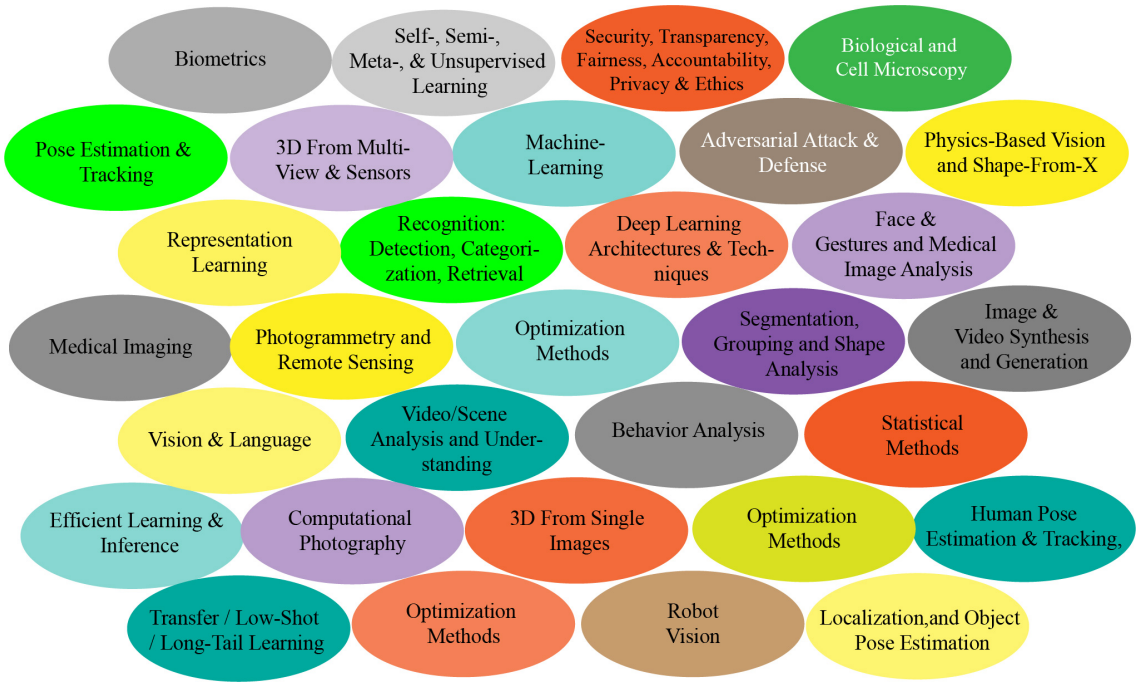


Figure 1.1. – Session on CVPR 2022 [VC22], the largest computer vision conference. Each of the sessions represents a research direction in computer vision.

ingful and discriminative features from visual data, as well as designing efficient algorithms for object detection, segmentation, and tracking, remain ongoing research objectives. Another significant challenge in computer vision research lies in handling large-scale datasets. With the advent of deep learning models, the demand for vast amounts of labeled data has increased exponentially. Collecting, annotating, and managing these datasets require substantial time, effort, and resources. Moreover, ensuring data diversity and avoiding biases present additional challenges. Researchers must address these issues to create comprehensive and representative datasets, fostering the development of robust vision algorithms.

The lack of interpretability and explainability in computer vision models is yet another critical challenge. Deep learning techniques, particularly convolutional neural networks (CNNs), have shown exceptional performance in various vision tasks. However, these models often operate as black boxes, making it challenging to understand their decision-making process. Interpretable models are crucial for applications such as medical diagnosis, autonomous systems, and legal contexts where transparency and trustworthiness are paramount. Furthermore, ethical considerations and biases embedded within computer vision systems have emerged as important challenges. Vision algorithms trained on biased datasets may inadvertently perpetuate social biases and discriminatory practices. Ensuring fairness, accountability, and transparency in vision algorithms, and addressing ethical concerns surrounding privacy and surveillance necessitates careful examination and development of appropriate guidelines and regulations. Finally, the use of image processing systems in real-world scenarios is a major challenge. The adaptation of image processing algorithms to different environmental conditions, real-time processing constraints, limited computational resources, and the need for energy efficiency are critical factors that must be considered. In addition, the robustness of image processing systems against external attacks and

their ability to handle unexpected situations, such as novel object classes or unforeseen scenarios, continue to be the subject of research.

In the following chapters, I address these and other important challenges in image processing research and highlight the complexities involved in developing advanced image processing systems. Exemplifying three applications, I also show possible novel solutions to problems that have so far hindered the applicability of computer vision in these applications. We discuss recent techniques, methods, and emerging trends that researchers are actively pursuing to overcome these obstacles. We demonstrate why standard methods cannot simply be applied to an existing problem when developing computer vision software. Each method must be customized to the specific application scenario in order to obtain a working solution. A scientific approach is required that clearly defines a framework for the particular application. Open-world applications, where objects or contexts can be recognized in arbitrary input images, are still too big a challenge to the current state of the art. However, as research moves towards general intelligence, such applications may become easier to implement in the near future. So far, however, algorithms and the available data still limit application possibilities and it is often uncertain which strategy to choose for a particular application. Therefore, there are a number of challenges in the practical development of applications that deal with high variant objects. In the following list are some identified practical issues that every developer has to face and find solutions for when developing applications with computer vision.

- How does the data need to be captured? Which sensor, depth, speed, format, framerate is needed?
- Are all labels clear or is there an undetected pattern in the data that needs to be discovered?
- Is there data that can be used as enrollment? Is it an identification, verification, segmentation, or other task?
- Is there a fixed number of classes between which the algorithm should distinguish?
- Is there enough annotated data that one can use for training?

In designing new CV applications, however, there are also some general challenges that cut across research areas, reflecting the complexity and intricacy of development.

1. Image Acquisition: Images acquired in real-world scenarios often suffer from various quality issues, such as noise, blur, low resolution, or lighting variations. Dealing with these challenges and developing algorithms that can handle such variations robustly is a significant hurdle.
2. Data Quality: Accurately detecting and recognizing objects in images or videos is a fundamental task in computer vision. However, this task becomes challenging due to factors such as occlusions, cluttered backgrounds, scale variations, and viewpoint changes. Developing algorithms that can handle these challenges and achieve high detection and recognition performance is a key area of research.
3. Availability of Data: Deep learning approaches, which have shown remarkable success in computer vision, often require large-scale annotated datasets for training. Collecting and annotating such datasets can be time-consuming, labor-intensive, and costly. Additionally, ensuring the diversity and representativeness of the data is crucial for developing robust models.
4. Computational Costs: Many computer vision applications such as robotics or autonomous vehicles require real-time visual data processing. Achieving high-speed and low-latency processing while maintaining accuracy is a significant challenge, especially when dealing with computationally demanding algorithms.
5. Handling Large Amounts of Data: As computer vision systems become more sophisticated, they often generate vast amounts of visual data. Efficiently storing, processing, and analyzing such data pose challenges in terms of storage requirements, computational resources, and data management.

6. Robustness to Adverse Conditions: Computer vision algorithms should ideally be robust to various adverse conditions, such as changes in lighting, weather conditions, or viewpoint. Ensuring that the models can handle such challenges and maintain performance in real-world scenarios is a significant concern.
7. Ethical Considerations and Bias: Computer vision systems can inadvertently inherit biases present in the training data, leading to unfair or discriminatory outcomes. Addressing these ethical considerations and developing methods to mitigate bias in algorithms is essential to ensure fairness and accountability.
8. Integration with Other Technologies: Computer vision applications often need to be integrated with other technologies, such as natural language processing, robotics, or augmented reality. Ensuring seamless integration and interoperability with these technologies can present integration challenges.
9. Privacy and Security: Computer vision systems often deal with sensitive visual data, raising concerns about privacy and security. Implementing appropriate measures to protect data privacy and ensure the security of computer vision applications is crucial.

The present work approaches these research challenges from several angles. Algorithmic-, data-, and application-level solutions are presented to ensure the applicability of the selected applications.

1.3. Applications

In this section, I will discuss the applications of computer vision in biometrics, as well as two additional applications: entrance gate control and textile defect detection. These applications are considered in this thesis, and the contributions aim to bring improvements in each respective area. The applications presented here are part of the current research on biometrics.

1.3.1. Biometrics

Biometrics is mainly used in use cases where individuals are to be identified or verified. Which biometric modality is used depends strongly on the respective requirements of the scenario. In particular, the parameters: uniqueness, constancy, measurability and universality determine the choice of a characteristic. In addition, there are many other characteristics that are currently part of the research to make biometrics better and more secure.

Face Recognition Biometric authentication is becoming increasingly popular. The goal of a biometric system is to automatically recognize individuals based on their biological and/or behavioral characteristics. In terms of convenience, it is required that people verify themselves in an automated process without human supervision. Commonly implemented and studied biometric modalities include: fingerprint [TBD*21], face [WD21], and handwriting [HS20]. Modalities are differentiated by their universality, uniqueness, constancy, measurability, performance, acceptability, and circumvention. Face recognition systems are especially increasingly used in our daily life. Inexpensive camera sensors and high accuracy and user-acceptance make face recognition one of the most commonly used kinds of biometric authentication. The level of performance that can be achieved with state-of-the-art methods depends on the capturing environment. Good performance can be achieved in controlled environments, e.g., in automated border-control. In that case, an image captured according to the ISO/IEC Standard 19794-5 [ISO05] and stored on a passport is used as a reference and compared with a probe image captured under controlled conditions at the border crossing. In that use case, both images are taken in controlled conditions. Following constraints on how a reference should appear, visible information such as gender, expression, pose, and eye color is discernible by an observer pertaining to the face. More recently, there has been an increased demand for recognition of unconstrained face images, such as those captured by mobile devices and surveillance cameras [BZD*15]. Compared to controlled environments, the recognition of

unconstrained face images is more difficult. The main challenges in face recognition are: degradation of face image quality and the wide variations of pose, illumination, expression, and occlusion that are often encountered in images [DT16]. In addition to access control, biometrics serves a diverse range of purposes with a primary focus on individual identification and verification. Contemporary mobile devices, including smartphones and tablets, incorporate biometric sensors like fingerprint scanners and facial recognition systems to facilitate secure device unlocking and authorization for mobile payment transactions. The utilization of biometrics enhances security and convenience in comparison to traditional PINs or passwords. In the realm of forensic investigations, biometric analysis assumes a critical role. Techniques such as fingerprint analysis, DNA profiling, and facial recognition are extensively employed to establish con- nections between individuals and crime scenes, facilitate suspect identification, and furnish evidentiary support in legal proceedings. Furthermore, biometrics plays a pivotal part in video surveillance systems, enabling the identification and tracking of individuals in public spaces. It contributes to security investigations, threat detection, and the proactive monitoring of high-security areas. Besides better algorithms, another way to increase the precision of face recognition is to use more than one modality at the same time. So-called multi-biometrics offers improved accuracy, robustness, security, handling of variations and increased uniqueness and flexibility compared to single modality systems. These advantages make multi-biometrics an attractive choice in various biometric recognition applications. Multi-biometric systems require more time for capturing and processing and often need more than one sensor. This is inconvenient, especially in scenarios without the cooperation of subjects. Commonly known modalities that can be captured at the same time, while capturing the face, are: iris and ear. Relatively unknown is the role of freckles or melamine face pigmentation (MFP) in this context (see Figure 1.2 for two samples of made visible MFP).

Figure 1.2. – Images taken in ultraviolet spectrum of the same person can make melamine pigmentation visible.

Face Presentation Attack Detection The accuracy of face verification systems has improved significantly since the advent of deep learning but their vulnerability to presentation attacks remains a major challenge. Presentation attacks are defined as “presentation to the biometric data capture subsystem with the goal of interfering with the operation of the biometric system.” [Int16]. They range from very simple low effort attacks like printed face images or replayed videos to more sophisticated attacks involving high quality disguises and masks (see Figure 1.3). Presentation attack detection (PAD) are approaches to prevent presentation attacks from a single image or series of images, using different properties like motion, texture, or life signs. There is currently no detection method that is absolutely safe because of growing concerns about the ability to resist spoofing attacks - specifically in scenarios such as automated border control and authentication on mobile phones. Software and hardware-based approaches have been developed to distinguish between real faces and so-called "presentation attacks" [RB17]. The goal of these attacks is to subvert the face recognition system by presenting a facial biometric artifact. Popular presentation attacks are: printed photos, replaying video using an electronic display, and 3D face masks. Three-dimensional masks and high-quality 3D-prints can especially very often overcome PAD. Commonly known methods include additional analysis of images captured in different wavelengths, especially in

the infrared (IR) and near-infrared (NIR). Their general vulnerability in practice is that 2D and 3D face images of people are commonly available and/or easy to shoot or can be generated even from a single image [ASJT17]. In order to increase the recognition performance and guarantee more robustness against attacks, biometric systems using more than one biometric modality at the same time are increasingly used (e.g., face and fingerprint). These multi-biometric systems require more time for capturing and processing and often need more than one sensor.

This is inconvenient, especially in scenarios without the cooperation of subjects. Better methods of detecting attack attempts, such as using properties in the hyper-spectral range that are invisible to humans, would improve the applicability of biometrics.

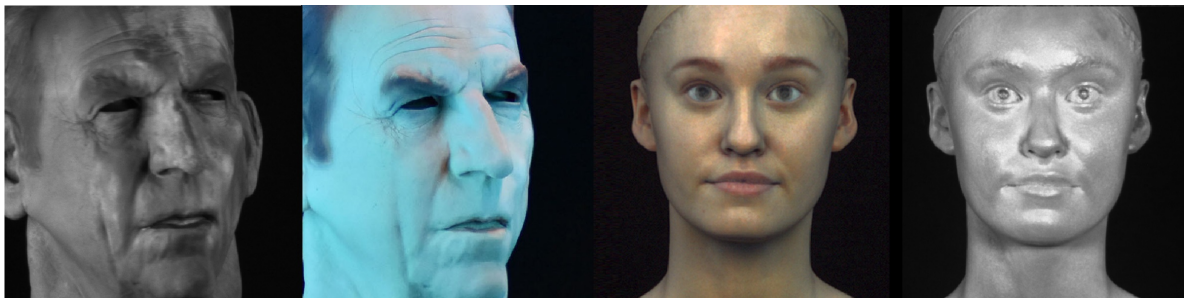


Figure 1.3. – Images of mask attacks, taken in ultraviolet and visible spectrum.

Handwriting Recognition Handwriting is a behavioral characteristic that is individual for every writer [ALSC02]. Therefore, handwriting identification has many applications, e.g., in security or forensics. One distinguishes between handwriting identification, which is a one-to-many comparison, and handwriting verification, that refers to a one-to-one comparison [WHL14]. There is offline writer identification, also known as static or static-dynamic writer identification, which involves analyzing static images of handwritten text or signatures. In this approach, the writing samples are acquired by scanning or capturing the handwritten documents or signatures in a non-real-time manner. A distinction is made between offline and online writer recognition. Offline writer identification, also known as static or static-dynamic writer identification, involves analyzing static images of handwritten text or signatures. In this approach, the writing samples are acquired by scanning or capturing the handwritten documents or signatures in a non-real-time manner. Online writer identification, also referred to as dynamic or dynamic-static writer identification, is based on the analysis of the dynamic features of the writing process. In this approach, the data is acquired in a real-time manner as the writer interacts with a digital device or a specialized digitizing tablet. Writer identification is utilized in forensic investigations, document authentication, cybersecurity and fraud detection, authorship attribution, plagiarism detection, psychological profiling, historical and archival research, and personalization/user verification. In forensic investigations, it aids in identifying individuals responsible for anonymous letters or forged documents. Document authentication relies on writer identification to establish the authenticity of historical manuscripts or legal papers. In cybersecurity, it helps detect impersonation attempts and malicious activities. Plagiarism detection relies on it to identify instances of copied content. Psychological profiling utilizes handwriting analysis for understanding personality traits or mental states. In historical research, it aids in identifying authors of historical documents. Lastly, writer identification enables personalization and user verification in systems or applications. A specific scenario for the application of handwriting identification is the analysis of questionnaires. In certain surveys each person is allowed to fill out one questionnaire only, so the developed application needs to ensure that this demand is fulfilled. If someone submits more than one questionnaire, this person is double enrolled which needs to be detected and removed. This is often the case when using anonymous offline questionnaires for reviewing services or products. In this

case, it is often not guaranteed that a reviewer does this only once as intended. Detecting and preventing double enrollment attacks is crucial to uphold the security and integrity of a biometric system.

Research Questions Classification performance is crucial in face recognition as it directly influences the system's ability to correctly identify individuals and is essential for the reliability, effectiveness, and trustworthiness of a face recognition system. Compared to systems with only one modality, multi-biometrics offers, among other things, higher accuracy, robustness, safety, and handling of deviations. It is assumed so far that MFP is not commonly used as a primary biometric modality due to its limited stability and variability. The artist Thomas Leveritt [Lev14] showed that image captures in the ultraviolet (UV) wavelength reveal hidden skin pigmentation. Even people that appear not to have any freckles sometimes show certain MFP in the UV spectrum (see Figure 3.1 I-III A/B). Melanin face pigmentation (MFP) is naturally caused and indicates cell damage. So far, it is unknown if this pigmentation can be used as a biometric feature for face recognition and/or as a modality. By answering the following research question, it can be found out whether the applicability of face recognition can be increased by MFP features.

> **Research Question 1:** Is melamine face pigmentation a biometric modality and can it improve conventional facial recognition algorithms?

The security of the commonly used face recognition algorithms is often doubted, as they appear vulnerable to presentation attacks. Detection of attempted attacks (PAD) is important for the acceptance of such systems and severely limits the applicability of biometrics. There are a number of detection methods while the exploration of skin properties like MFP in the ultraviolet spectrum has not been studied. Additionally, double enrollment presentation attacks are hindering the trustfulness of handwritten text-like surveys. Therefore, the following research question is drawn to tackle presentation attack problems:

> **Research Question 2:** Are there novel feature types or modalities that can be used to increase PAD performance?

I have examined features in two publications (see Figure 1.6) that have received less attention in research to date. The conclusions to answer this research question are based on the research shown therein.

1.3.2. Entrance Control

Systems for the safe separation of individuals are of high importance at all kinds of security zone entrance. They are used at access points of critical infrastructure, public transportation, event locations as well as in high security business and military areas. Autonomous access control gates are being used more and more often in accessing areas with a high security level. People with permission pass through this designated transit space to access a secured area. The foremost advantage of these systems is robustness against social engineering - meaning that an authorized person takes an unauthorized person into the secured area (tailgating).

One example is a passenger passing the entrance gates without paying in metro stations. This kind of behavior is called fare evasion, and it is troublesome and costly to prevent. As a typical type of fare evasion, tailgating refers to following a fare-paying passenger through the gate. It can be dangerous because the passenger risks being injured by the barrier gate. To detect tailgating fare evasions automatically, the existing surveillance cameras in stations can be utilized to provide a visual-based method at a low cost and efficiently. However, occlusion by crowds during rush hours can lower the accuracy of regular recognition methods based on convolutional neural networks. Moreover, the behavior of a tailgater can be similar to that of other fare-paying passengers. Another example where separation of individuals is needed are so-called mantrap portals (see Figure 4.1a). They provide a closed area with two doors - one as an entrance and another for leaving this area. Permitted subjects enter and close the portal, so that software can verify that only one subject is present in the transit space. To authorize them,

Figure 1.4. – Piggybacking attack (left) Bona fide authentication of mantrap portal (right)

biometric information, PINs, or passwords are used. After successful verification, the system unlocks the final door to give access to the secured area. A general problem is that an authorized person can take an unauthorized person into the secured area. This practice is entitled as “tailgating” or “piggybacking” (see Figure 1.4 left). Many barriers like (drop-arm-) turnstiles are therefore equipped with sensors like infrared break-beams in order to eliminate this vulnerability. However, these available systems are designed to achieve high flow rates and can easily be defeated. In places where higher security is needed, mantrap portals are used. These systems regulate the access of only a single person through a transit space. Permitted subjects enter and close the portal, so that a software can verify the number of people present in the transit space. After successful verification, the system unlocks a second door to give access to the secured area. Previous research in the field of tailgating/piggybacking prevention shows that none of the existing systems are completely safe. Computer vision approaches based on video captures could detect intrusion, analyzing movements and distinguishing it from certain behavior. We have therefore developed various approaches for detecting this behavior on the basis of video or single-frame recordings. To this end, I first developed an image database, on the basis of which further investigations were carried out and which will be presented in later sections. One approach is the analysis of infra-red thermal images, another approach is the analysis of motion in the portal using optical flow. Also 3D depth information and capacitive sensor technology were investigated and each represent new contributions to the solution of this problem (see Figure 1.6), which are published in this thesis.

Research Question In this area, there is a plethora of algorithms that seem potentially suitable. One disadvantage of all these single-sensor-based methods is the creativity of the attackers to always invent new attempts to overcome them. The camera angle is especially hindering as it allows people to hide on the floor or between the legs of a permitted person. Sensors embedded in the floor can be of great help here but are limited by the fact that all feet must be on the ground. Therefore, the following research question arises that needs to be answered in order to enable autonomous mantrap portals.

> **Research Question 3:** Which technologies are suitable for autonomously detecting piggybacking and tailgating when accessing restricted areas?

In answering this question, there are several practical factors to consider that may increase or decrease applicability. These include, for example, the carrying of objects, the obligatory cooperation of the user, e.g. to position himself in certain places around the room.

1.3.3. Textile Defect Detection

Fiber defect detection is one of the tasks that is recently gaining a lot of attention from textile producing and textile laundry industries. As quality, environmental awareness, and cost reduction are increasingly important factors in the use of cleaning textiles, more and more reusable textiles are being used. For industrial textile users, the quality of the fabric as well as the cleanliness and undamaged weaving of the textiles are some of the most important factors in the selection of a textile supplier. The sorting of textiles or fabrics according to defects or product categories plays an important role for companies operating in textile cleaning, textile recycling, or returns handling. Sorting used textiles according to defects or quality standards is price relevant. A defect can occur in the case of the regular structure of the material surface being disturbed, e.g., through holes, cuts, or stains. There are about 70 different types of defects in the literature [Cou00]. Automation processes are known to be used for textile sorting and washing. The textiles are often transported in voluminous shape on a conveyor belt and visually inspected by an employee. Nowadays more than one billion cleaning textiles in Europe are being leased and reused per year. Besides the big volume of processed pieces, quality assurance of used industrial textiles has remained a predominantly manually operated task. Compared to humans, automated systems can have several advantages, such as lower costs and higher reliability and accuracy. Quality assurance after cleaning is a cost intensive operating process. Lowering its costs will lead to an overall cost reduction and may therefore encourage more customers to start using reusable industrial textiles. With the increasing performance of artificial intelligence, automatic textile defect detection has become one of the most relevant areas in this domain. So far, recent work in the field of textile inspection deals mostly with continuous 2D textures (spread-out). This is because textile inspection algorithms are mostly used during furling in the production process. These applications need a solution to be used by the cleaning industry that handles textiles individually in an assembly-line workflow (see Figure 1.5 left). Due to high flow rates of textiles, an automatized mechanism to spread out textiles mechanically while in movement has not yet been invented. Therefore, a focus on inspection of defects in a pile-like arrangement, where every item is still dealt with separately, on an assembly-line, is needed. The uneven surface, varying colors of sewing pattern, and weaving of different textile fibers make these cleaning textiles highly variant objects. Furthermore, textiles differ in the composition of fibers which include cotton, linen, polyester, or compositions. Previous research on outspread fabrics achieves high recognition rates but is not resistant against some effects caused by voluminous shapes.

Figure 1.5. – (left) Quality Assurance of Textiles served Individually. (right) Interface of a defect detection system.

Shapes, folds, edges, borders, overlapping edges, and ambient occlusion are some of these effects. They have a negative impact on the correct detection of fiber defects using conventional methods. The voluminous shape of textiles also results in a loss of focus in some parts of the image.

Automated methods are known for the image-based detection of defects on textiles that are spread out, but these produce low detection rates when used on non-spread-out textiles. Spreading out of a textile, in particular a mechanical spreading, decreases the time needed for a computer-vision-based inspection and increases inspection quality. Difficulties affecting the image-based classification of textiles in general occur if:

1. Textiles are in motion
2. At least half of the surface is covered, since the textile is classified horizontally
3. Folds and shadows dynamically change parts of the surface
4. Textiles deform freely
5. Different patterns make the detection of defects difficult, in particular if a defect is a wanted defect

Also known are methods which deal with error detection in 2-D images of statically arranged textiles. However, these methods only partially solve the problem because:

- Textiles must not move
- Large parts of a lying textile are not considered (because they lie below on a non-imaged side of the textile)
- The detection performance of deformable surfaces, such as textiles (for reason 1) is too low

In this work, I therefore provide new contributions that increase the applicability of textile defect detection and extend it to non-baked fabrics. To investigate this use case in detail, I present a new database showing these textiles in different imaging situations. Several new methods will then be presented to enable recognition performance to a super-human level.

Research Questions There are a number of algorithms that have shown good results on data sets with spread out textiles in this area. Existing publications are trying textile defect detection using spectral, model-based, or statistical approaches. The biggest problem that all algorithms face is achieving a high level of generalizability while still maintaining a high recognition rate. Our experiments have shown that these approaches cannot be applied to voluminous cloths. The generalization of the existing algorithms to apply defect detection to cloths that are in the form of piles has not been sufficiently considered so far. Research is therefore necessary to develop new approaches to improve the transferability of the algorithms to non-spread cloth. For this reason, following research question in this field of study arises.

> **Research Question 4:** Is an algorithm able to generalize enough to detect defects on different materials?

To ensure real-time recognition, it is necessary that the processing steps are as fast and simple as possible. Visual quality control performed by humans is superior to computers in many respects, for example, a human can view an object from different angles. Humans also have a very low false positive rate because their eyes are a very reliable source. From the point of view of a possible automation, the question arises how an optimal acquisition scenario could look like, which not only represents a competitive setup from the point of view of recognition performance, but also under consideration of other parameters such as speed and costs.

> **Research Question 5:** Which system setup in combination with which algorithms allows an evaluation in a comparable time compared to humans?

Practical considerations such as efficient sorting and lighting must also be taken into account. Such factors also ultimately influence the applicability of the algorithms and thus the success of the research.

1.4. Structure of this Work

After providing the rationale and introducing the key aspects of this thesis, this section proceeds to offer an outline of the subsequent sections in the following manner: Chapter 2, Related Work. Specifies the relevant literature. It is grouped into the three application categories. The first section gives a background on biometrics including relevant historical work. The main components of biometric recognition and the PAD system are presented. The Sections 2.1.1-2.1.3 give an overview of the research areas: face recognition, presentation attack detection, and handwriting recognition. The second section presents the application of an autonomous entrance portal and gives an insight into the state of the art in related work in video analytics and people counting. Section 2.3 gives an overview of textile defect detection techniques and outlines the previously identified applications. The summary Section 2.4 gives a short overview of the related work in all three applications areas.

Chapter 3, Novel Methods within Biometrics. Biometrics is a science whose application in industry places particular challenges on algorithms. Here, I present various methods of solving practical requirements with scientific methods. Different novel algorithmic classification and segmentation methods as well as a novel database are presented. This includes methods for pre-processing, sensor-based, segmentation, normalization, and classification. This chapter is based on published papers [Sie14, SED16, SSG*18, SKM*20].

Chapter 4, Methods for Enabling Autonomous Entrance Control. Entrance control is still a largely manual activity in which people undertake a visual inspection. In order to replace this process with computer vision, various technologies and methods are presented in this chapter which can aid the implementation of an autonomous entry portal. This includes pre-processing, sensor-based, segmentation, normalization, and classification approach methods. This chapter is based on published papers [SWB16, SHK16, SFS*16, SDF*18, STvW*19].

Chapter 5, Novel Classification and Normalization Methods for Industrial Inspection of Textiles. Industry has a high demand for solutions that enable the visual inspection of highly complex materials. Classification and segmentation methods are presented that tackle these specific challenges. The results are given according to their application. This chapter is based on published papers [SBK16, SKH16, SSF*17, SPKK18, SFJG*21]

Chapter 6, Conclusions and Future Work. Conclusions and Future Work. This chapter recapitulates and concludes with a final view on the research questions of this dissertation and introduces potential future areas of research.

1.5. Summary

This chapter first presented a motivation leading to a set of research challenges. The listed challenges aim to improve the applicability of image classification methods for highly variant objects in applications. The research questions were based on the three applications presented: biometrics, autonomous input inspection, and textile defect detection, as shown in Figure 1.6. The first focus of the presented research questions dealt with the biometrics application and, in particular, improving the accuracy of facial recognition and preventing fraud in the form of PAD and "double enrollment". An in-depth analysis was performed and the use of hyperspectral imagery and low-level features was investigated. The second application area deals with the application of autonomous entrance portals for which different algorithms and sensor technologies have been studied. The third group of research questions aimed at applicability of defect detection on textiles when they are not in a spread-out state. Amounts on the algorithmic, data, and application side were created to improve the applicability of CV systems for the classification of objects with high variance in these scenarios.

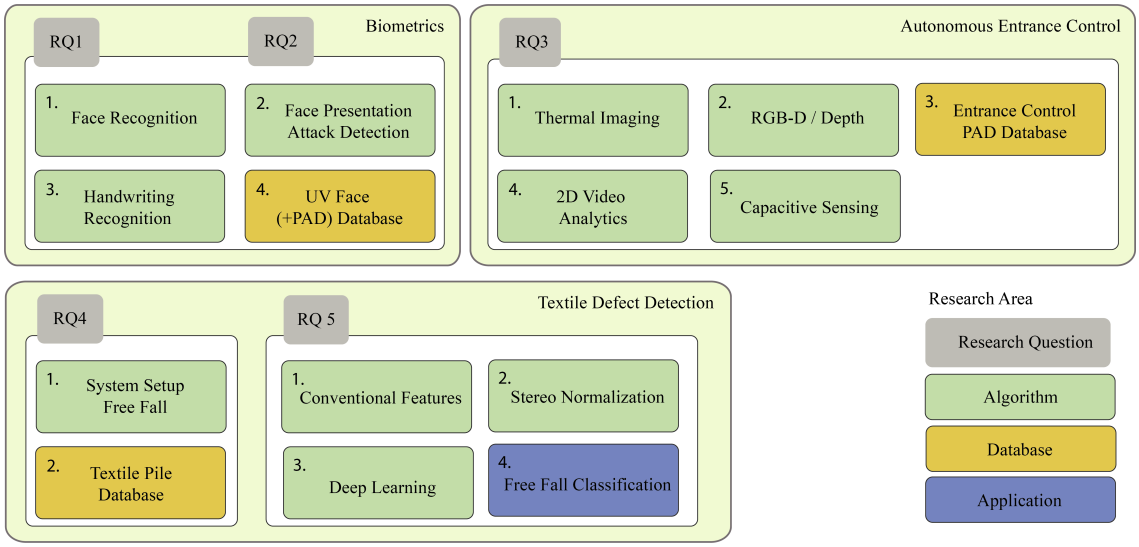


Figure 1.6. – An overview of the contributions in relation to the research questions posed in this thesis. Biometrics focuses on analyzing and enhancing the applicability of biometrics as an application in practice, autonomous entrance system describes the feasibility of such entrance systems using computer vison, and textile defect detection provides a solution to boost the generalizability of algorithms for textiles in a pile-like structure.

2. Related Work

The primary objective of this chapter is to provide a thorough examination of the existing literature and state-of-the-art approaches in the domain of computer vision. By critically analyzing previous research, we aim to identify gaps, challenges, and opportunities for further investigation. Additionally, this review serves as a foundation for our proposed methodology and research contributions, offering insights and inspirations from prior work. It describes the state of the art of the three applications - biometrics, entrance portals, and textile defect detection.

2.1. Biometrics

The following sections will provide an overview of the related work in two specific areas within the broader field of biometrics: face recognition and presentation attack detection in face and handwriting identification. These areas have been extensively researched, and understanding the existing literature is crucial for building upon the current knowledge and addressing the research questions in this thesis.

2.1.1. Face Recognition

Developing biometric applications, like facial recognition, has become a significant objective in the creation of smart cities. With this in mind, engineers and scientists worldwide are working on developing robust and precise algorithms that can be utilized in daily life. While passwords remain the most commonly used recognition method, all security systems must protect personal data. Thanks to advancements in information technologies and security algorithms, biometric factors have become increasingly popular for recognition tasks [FAKD22, TC17, GBDR*22, BDKK22]. By identifying individuals through physiological or behavioral characteristics, biometric factors offer several advantages, such as the ability to recognize a person simply by their presence in front of the sensor, eliminating the need to remember multiple passwords or confidential codes. Consequently, recognition systems based on a range of biometric factors, such as iris [BDR*20], fingerprints [TBD*21], handwriting, and face, have been implemented in recent years. In the realm of smart cities, developing biometric applications like facial recognition has become an essential objective. To achieve this goal, scientists and engineers worldwide are concentrating their efforts on creating more robust and accurate algorithms and methods that can be implemented in daily life. It is crucial for all security systems to safeguard personal data, and while passwords are the most commonly used recognition method, information technologies and security algorithms have enabled the use of multiple biometric factors for recognition tasks [FAKD22, TC17, GBDR*22, BDKK22]. These biometric factors allow for the identification of individuals based on their physiological or behavioral characteristics and offer numerous advantages, such as the ability to recognize a person by merely by being present in front of the sensor, eliminating the need to remember multiple passwords or confidential codes. Before delving into the techniques used, it is important to briefly describe the challenges that must be overcome in order to perform face recognition accurately. The key features that make a face recognition system effective include its ability to process both images and videos, function in real time, be robust in varying lighting conditions, be independent of physical characteristics such as hair, ethnicity, or gender, and work with faces from various angles. Different types of sensors such as RGB, depth, EEG, thermal, and wearable inertial sensors are utilized to collect data.

These sensors provide additional information and assist face recognition systems in identifying images of faces in both static images and video sequences.

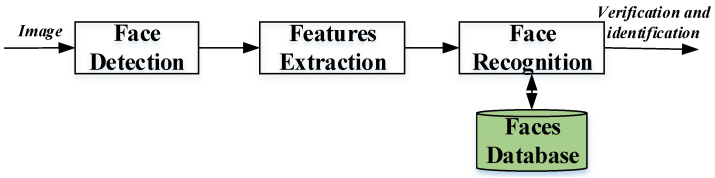


Figure 2.1. – Face recognition follows usually the process of detection-feature estraction and recognition by using a database of faces. More details can be found in the work of Kortli et al. [KJAFA20]

The development of a robust face recognition system involves three basic steps - face detection, feature extraction, and face recognition (see Figure 2.1). Face detection is used to locate human faces in an image, while feature extraction extracts prominent features of the face image to identify it. It is responsible for extracting a set of feature vectors that describe prominent features of a human face, such as the mouth, nose, and eyes. Each face is characterized by its size, shape, and structure, allowing it to be identified. Finally, the face recognition step involves comparing the features of the detected face with known faces in a specific database to make the acceptance or rejection decision. It involves comparing the features extracted from a face image with known faces stored in a database for two general applications - identification and verification. In the identification step, a test face is compared with a set of faces to find the most likely match. In the verification step, a test face is compared with a known face in the database to make an acceptance or rejection decision. Several techniques such as convolutional neural network [TIH*23], and k-nearest neighbor [KBG17] are used to address this task effectively. The natural differences between individuals make using facial characteristics for biometric recognition attractive due to the clear distinction between different individuals. Traditional solutions for face recognition rely on handcrafted features such as texture, keypoints, and descriptors. However, recent advances in deep learning and the availability of massive training datasets have led to breakthroughs in face recognition, even with low-quality and unconstrained data samples [WD21, GZ19].

Face recognition algorithms have significantly advanced in the last few years through the emergence of new methods like robust quality assessment [HOGF*19, LT21] and state-of-the-art algorithms [DGXZ19, TIH*23] However, their practical use in unconstrained environments and concerns about spoofing still remain a challenge, especially in big datasets. Images in other spectra therefore show potential to improve the performance and reliability of face recognition and PAD as they provide more information which is undetectable in the visual band (see Figure 2.2). It appears that the focus in multi-spectral face recognition lies currently on the infrared spectrum as it has some advantages under difficult conditions, such as dim light or foggy/dusty environments. Some of the later studies propose multi-spectral and hyper-spectral system processing as the consequent way to achieve even better results in multi-spectral face recognition [NB15, BKR*10, CSB20].

While facial images in near-infrared or short-wave infrared spectrums have been widely studied in research and in practice, facial images in the ultraviolet spectrum have been rather poorly studied, which is why only few publicly available databases exist [FMPZD18]. The UV spectrum in particular is yet to be examined in the context of face recognition, even though several studies mention its potential for this use [Bou16, BH16]. Observation of the human skin in the UV spectrum was first examined in the medical field rather than in the context of biometrics. Cooksey et al. [CA13] measured the reflectance of human skin in specific spectral bands from 250nm (UV) to 2500nm (IR). Narang et al. [NBH15] compared face recognition algorithms for images in UV vs. UV and UV vs. VIS. It has been recognized that UV light can detect pigment changes and skin damage which are not noticeable in the VIS light spectrum [Ful97]. In UV light, these characteristic pigment changes

Figure 2.2. – Random sample images under different illumination conditions taken for one subject from the MSpectra database. (Spectra from left to right: VIS-Nat, VIS-Fluo, NIR-Nat, NIR-Ifr, UV-Nat, UV-Fluo) [FMPZD18].

usually appear as dark spots on the skin. In Section 3.1, we explore the benefits of using UV images for face recognition with a focus on the MFP.

2.1.2. Face Presentation Attack Detection

In Face Recognition (FR), physical and behavioral traits are used for secure and automated authentication. However, vulnerabilities still affect FR systems, as they are exposed to various attacks. These include direct and indirect attacks like disguise or makeup [RDB21], and presentation attacks (PAs). Via PAs, an imposter can be authenticated as a genuine user by presenting fake facial attributes, using tools such as photos, videos, and masks [SOPP18]. Another example of a PA is face morphing attacks. They aim to create face images that are verifiable to be the face of multiple identities, which can lead to building faulty identity links in operations like border checks. This was demonstrated by Ferrara et al. [MAD14], who found that a single attack image can match more than one person, even with the assistance of human experts. If used with travel or identity documents, these attacks can enable multiple individuals to authenticate themselves based on the document's alphanumeric information, leading to potential illegal activities such as human trafficking, illegal immigration, and financial

fraud. Attacks introduce distortions to the sensor output images, which provide cues for anti-spoofing methods to detect fake images. Presentation attack detection (PAD) identifies whether the image is genuine or fake (see Figure 2.3. If a spoofing attack is detected, access is rejected, and if no attack is detected, the system processes the image for authentication, and access is either granted or denied.

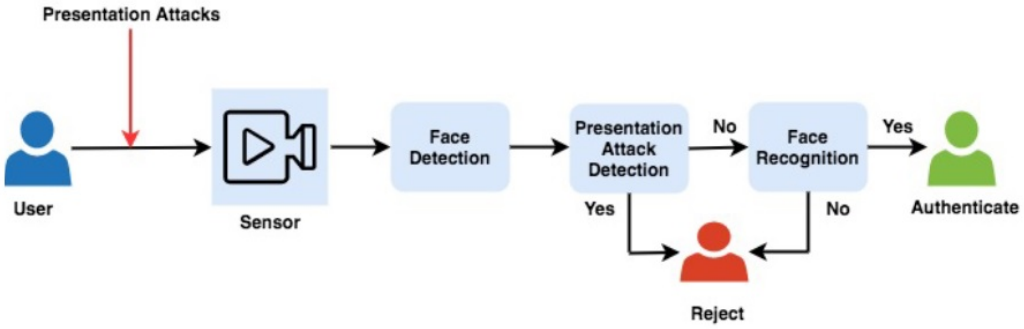


Figure 2.3. – Face Recognition System with Presentation Attack Detection [AEJ21].

Active imposter presentation attack detection algorithms can be categorized into hardware- and software-based. Software-based algorithms are cheaper, space saving, and include static and dynamic approaches. Static PAD algorithms use single images while dynamic ones also capture temporal information. They can analyze micro- textural patterns [RB17] and/or motion [DMNRD12] but mostly fail when a trained model is used in a different environment or on other datasets. Damer et al. [DD16] reported good results in a motion magnification-based approach using histograms of oriented optical flow. A limitation of this approach is the human physiological rhythm itself and computational costs. Hardware-based multi-spectral algorithms analyze several images in distinct regions of the electro-magnetic spectrum individually [RRV*17]. There are also multi-sensor approaches, where multiple spectral bands are used by different sensors at the same time. The spectral band can be divided into VIS [400nm – 700nm], IR [780nm-15 µm], NIR and short-wave (SWIR) bands. Medium-wave and long-wave are used due to being "thermal". A multi-spectral PAD approach was presented by Ramachandra et al. [RB17], where they used seven bands between 425 and 930nm. Presentation attacks were performed with images printed on laser and ink-jet printers. Almost all bands were vulnerable to attacks with laser printed images, and vulnerable to presentation attacks with images printed on an inkjet printer as well. In other words, IR-imaging alone does not significantly improve the quality of a PAD system. Multi-sensor/cross model approaches can take advantage of the different reflection properties of material in a different spectrum. In other words, knowing that human skin reflects IR light quite differently than, e.g., silicon, allows to detect presentation attacks by comparing images which capture both spectra. The effectiveness of this method is demonstrated by the known FaceID, used by Apple iPhones. Apple uses multiple face-depth maps and high-resolution images in the IR spectrum to generate a template. But while active IR or NIR images show advantages especially in robustness to illumination and exhibit special characteristics of the human skin, this method has been spoofed as well by using a 3D mask. Steiner et al. [SKJ16] presented such an approach and evaluated it on a database of SWIR and VIS images including attacks using (partial) disguises and masks. Their methods show almost perfect spoof detection (1% in case of disguises) while having an FRR of around 5%. Due to the MFP ascribed properties, we think that these features should also be useful for PAD. Despite intensive research, however, we could not find such an approach in previously published literature..

2.1.3. Handwriting Writer Recognition

Handwriting is a behavioral characteristic that is individual for every writer [ALSC02]. Therefore, handwriting identification has many applications, e.g., in security applications or forensics. One distinguishes between handwriting identification, which is a one-to-many comparison, and handwriting verification, that refers to a one-to-one comparison [WHL14].

At first, writer identification used only handcrafted features, but with the introduction of deep learning, new methods were proposed. Prior to these deep learning methods, various classifiers such as SVM, K-NN, and neural network were utilized in combination with tools like PCA and LDA to enhance the discriminatory ability of handcrafted features. In the next two subsections, we will briefly discuss handcrafted features and deep-learning-based approaches for writer identification. Various techniques have been used to exploit the differences in visual shapes of handwritten characters for writer identification. For instance, Schomaker and Bulacu utilized connected-component contours and its probability density function [SB04], while He et al. used a hidden Markov tree (HMT) in the wavelet domain. Other approaches include the development of a continuous character prototype distribution feature extraction technique by Tan et al. [TVGK09] and the use of K adjacent segments (KAS) by Jain and Doermann to model character contours [JD11]. The latter also used contour gradients and pseudo alphabets as features. Meanwhile, He et al. extracted features such as junction detection, final junction refinement quill, and hinge, and linked it with a learned codebook [HWS15].

In recent times, there has been an increased interest in deep learning as convolutional neural networks (CNNs) have demonstrated their effectiveness in extracting distinguishing features from handwritten texts. DeepWriter [XQ16] utilized multi-stream CNNs to acquire various representations of text images. Rehman et al. augmented text images with various techniques and generated multiple patches which were fed into an architecture similar to AlexNet for feature extraction. These features were then classified using a support vector machine. Keglevic et al. [KFS18] designed a triplet network to compute the similarity measure between different patches and trained it by maximizing inter-class distance and minimizing intra-class distance. Global features of the document were then obtained by aggregating the vector of local image patch descriptors. Nguyen et al. [NNI*19] generated tuples of text images by randomly sampling characters as input for their CNNs. They trained the CNNs to extract sub-region, character, and global level features and aggregated them effectively to predict the identity of the writer. He et al. [HS20] designed FragNet which builds a global feature pyramid first and then a local fragment pathway that leverages fragments of global feature pyramids to make separate writer identity predictions for each writer. Javidi and Jampour [JJ20] quantified the thickness of handwritten documents using handwriting thickness descriptors (HTD). Resnet-18 was used to extract features from the text images, which were combined with HTDs for classification. Srivastava et al. [SCP22] proposes three different deep learning models that use different architecture-based components suitable for identifying and capturing various aspects of a writer's technique and style.

A specific scenario for the application of handwriting identification is the analysis of questionnaires. In certain surveys, each person is allowed to fill out one questionnaire only, so the developed application needs to ensure that this demand is fulfilled. If someone submits more than one questionnaire, this person is double enrolled, which needs to be detected and removed. To our knowledge, a database that yields complete questionnaires does not yet exist. Known handwriting databases like CEDAR [SV10], NIST [WGJ*92] and CENPARMI [SNL*92] contain mostly isolated characters or single words that do not reflect unconstrained environments. The databases IAM [MB02], ICDAR 2013 [LGSP13] and the CVL dataset [KFDS13] contain large amounts of unconstrained handwritten English sentences but are limited by a small amount of writers ($<$350). Therefore, a novel database was created that contains handwritten texts and rating parts from actual review forms. From the samples of this new database different features were extracted. Most of the features get extracted from single words or lines, so a good segmentation process needs to be employed [CMS14]. One way of doing this is described by Schomaker

et al. [SB04] who used connected components as basic structure. There has already been a lot of research regarding the extraction of texture-based and allographic features [AMH09, MMB01a]. This work deals with the extraction of nine features of different types. To improve the outcome, all of these features were combined in a fusion process introduced by Damer et al. [DON14a]. The results show the impact of the different features onto the identification process and how this can be used to identify double enrolled handwriting.

2.2. Autonomous Entrance Control

Tailgating is a social engineering attack challenging physical security within organizations. It gained public traction in the year 1999 and has since remained a major concern in the field of security. While security guards could be placed at all entry points of a building to detect tailgating events visually, that would be very expensive for any sizable building. If it was possible to detect unauthorized entries automatically through computer vision, the guards could be alarmed to the gate when needed [Tuo19]. This led to the development of several anti-tailgating solutions [Dan20]. These solutions began with simple mechanisms like mechanical turnstiles, revolving doors, and mantrap systems and evolved into more modern technologies using infrared beams, 3D machine vision, face detection, BMI and face recognition combination, and an embedded solution using IP camera and video analytics. A critical analysis of these solutions uncovered certain weaknesses which run through most of them. These are the inability to detect two people side by side and the incapability of detecting multiple entries after a single access authorization [AC21]. Commercial systems that try to prevent tailgating and/or piggybacking, combine physical barriers like drop-arm turnstiles with sensors. But they must allow a certain range of weight or require an identity claim to pass the system, therefore they are not flexible to variations. Infrared beams, mounted at waist level, can easily be overcome by jumping or walking closely together.

Lower security level areas and/or where enforcing one-way traffic is needed, however, prefer to use turnstiles or access gates. Turnstiles allow only one person at a time to pass the secured area by using a ratchet mechanism. Disadvantages of this type of separation are its inconvenience for handicapped people and unauthorized tailgating. Optical turnstiles try to overcome these problems by using infrared beams to count individuals. Their mechanic functionality is easy to overcome, yet unaesthetic and inconvenient for some people. Therefore, optical drop-arm turnstiles (see Figure 2.4) are increasingly used. They allow individuals and handicapped people to pass with less delay than all other methods [McG13]. Larger cabinets allow people to carry trolleys, bags, and luggage. But even these systems are easy to overcome by tailgating or piggybacking as our review of the latest achievements shows. Tailgating is a vulnerability that can result in theft, property damage, and other unauthorized activity that is harmful to the system. In a study about a railway station, it was found that more than 1% of 298,799 entries (3,999) were instances of tailgating [CTC*19].

Manufacturers of mantrap portals [Boo16, Amr16] use scales [Soe16] or photo sensors [SIC16] to measure humans' weight or body shape. Systems using weight suffer from the principal disadvantage of a high tolerance in order to guarantee that subjects can pass with heavy objects. As the walls of such systems are often not covered by the scale, people might also be able to spoof these systems. In research, several computer vision methods have been developed in order to overcome the weaknesses of non-smart entrance gates. Most methods use a top-view perspective and pattern recognition methods to distinguish between one and more than one person in an observed area. Nevertheless, there are only a few academic studies on this topic. Hung et al. [HSC*22] proposes to use videos as input to estimate human pose data in each frame. Obtained incomplete skeletons are retained and multiple persons appearing in adjacent frames are matched, after which a sequence of skeleton data is generated for each pedestrian. Thirdly, a time series of the positional relationship between passengers and the gate is extracted and the passing interval of passengers is defined as the indicator for detecting tailgating. In their experiments, they showed showed that tailgaters could be distinguished effectively from others using

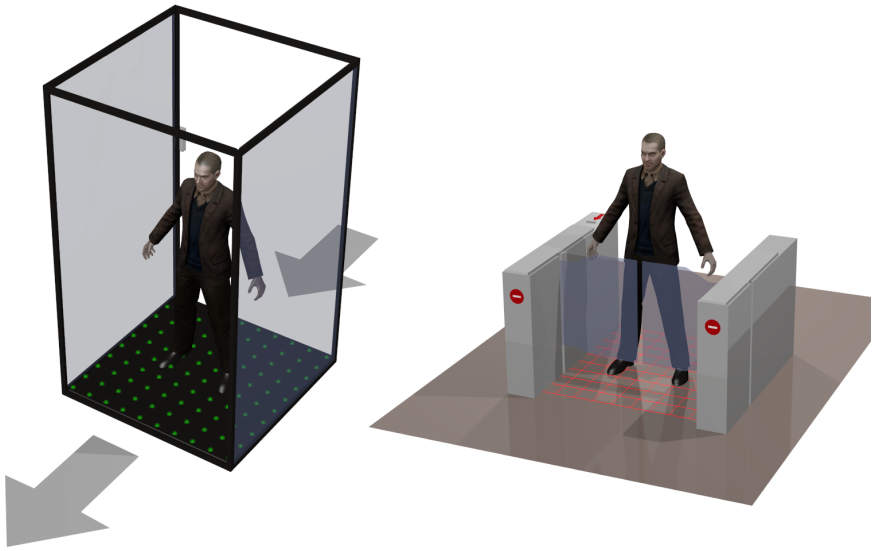


Figure 2.4. – a) Mantrap Portal b) Drop-Down Turnstile

a time series (see Figure 2.5) . Advances in computer vision have allowed video surveillance to be further automated to detect and classify motion trajectories [HMOS12]. In particular, such algorithms can be used in tandem with radio-frequency identification-based (RFID) technologies to detect tailgating [HCK*17]. Indoor location systems track users by using networks of beacons on access doors and RFID tags [TCCGRP*18]. The limitation of these technologies is that they require additional installations of physical equipment (like antennas) and distribution of tracking devices. They also introduce privacy and security concerns and can be overcome by people that know that this technology is being used [CTC*19]. A low-cost solution to this problem of tailgating was presented by using a single internet protocol camera and an embedded-based control unit combining with video analytic technology from a surveillance camera. The implementation of a BeagleBoard xM was combined with video analytics and three-stage checking. The system is able to achieve an accuracy rate of 90.8% with the frame size of 320 x 240 running at 7 FPS. Some other authors concluded that multi-modal data information could improve their results, e.g., by using other sensor data or establishing pose and position guidelines. Similar to the aforementioned use case, the patent [SFN05] describes a technique which provides increased levels of security for a mantrap portal through the use of two zones. Primary and secondary sensors are used in combination for continuous monitoring and detection of presence of unauthorized subjects prior to granting access to a secured area.

Unlike the use case in which a portal is used, there is some work focusing on the detection or counting of humans from different viewing directions. From in-the-wild video data there are anomaly detection methods trying to detect abnormal behavior in surveillance videos. This is a challenge task, because of the diversity of possible events. Nguyen and Meunier propose a single camera approach using a deep convolutional neural network (CNN) and learning a correspondence between common object appearances (e.g., pedestrian, background, tree, etc.) and their associated motions [NM19]. Their model is designed as a combination of a reconstruction network and an image translation model that share the same encoder. The former sub-network determines the most significant structures that appear in video frames and the latter attempts to associate motion templates to such structures. The training stage is performed using only videos of normal events and the model is then capable of estimating frame-level scores for an unknown input. The experiments on six benchmark datasets demonstrate

the competitive performance of the proposed approach with respect to state-of- the-art methods. Another approach that could also work on behavior identification is to minimize the energy of a human motion model that includes multiple personal, social, and environmental (PSE) constraints across cameras. The idea is to assign a 1–1 correspondence while modeling the PSE constraints. The authors optimize their model using a greedy local neighborhood search algorithm to restrict the search space of hypotheses and propose good results on the PRID and Grand Central datasets [MAIS16]. At first glance, research into the re-identification of people may give a further boost towards the efficient prevention of tailgating but this has been widely studied as a specific person retrieval problem across non-overlapping cameras while tailgating detection is independent from the identity of the person. Nevertheless, research on re-identification shares the same challenges that research on tailgating detection suffers from. The widely studied closed-world setting is usually applied under various research-oriented assumptions and has achieved inspiring success using deep learning techniques on a number of datasets. With the performance saturation in a closed-world setting, the research focus for person Re-ID has recently shifted to the open-world setting, facing more challenging issues. This setting is closer to practical applications under specific scenarios [YSL*21] but the achieved results are still far from being accurate enough to be usable for detecting reliable tailgating behavior. Another similar challenge is counting people on video data which could theoretically be used to count the number of people present in a certain space. Sun et al. propose a method for counting people in real-world cluttered scenes related to public transportation using depth videos [SAS*19]. The proposed method computes a point cloud from the depth video frame and re-projects it onto the ground plane to normalize the depth information. The resulting depth image is analyzed for identifying potential human heads. The human head proposals are meticulously refined using a 3D human model. The proposals in each frame of the continuous video stream are tracked to trace their trajectories. The trajectories are again refined to ascertain reliable counting. People are eventually counted by accumulating the head trajectories leaving the scene. Their accuracy varies between 75% and 93% which is a good result for video analysis of surveillance cameras but that might not be enough for being used in high security areas.

Automated video surveillance addresses people's real-time observation to describe their behaviors and interactions. In another publication, a multi-person tracking system for crowd counting and normal/ abnormal event detection in indoor/outdoor surveillance environments have been proposed [SJK19]. The presented system consists of four modules: people detection, head-torso template extraction, tracking, and crowd cluster analysis. Firstly, the system extracts human silhouettes. Secondly, people are detected by their head and torso as this is less varied and hardly occluded. Thirdly, each person is tracked through consecutive frames. Finally, the template marking is used for crowd counting, cued through localization and clustered via Gaussian mapping for normal/abnormal event detection. The experimental results on two challenging datasets of video surveillance such as PETS2009 and UMN crowd analysis datasets demonstrate that the proposed system provides 88.7% and 95.5% in terms of counting accuracy and detection rate. Only looking at the people counting task using 2D images, Pervaiz et al. achieved state-of-the-art results using a self-organizing map (SOM) to cluster particle flow [PGG*21]. Random particles are distributed, and features are extracted. By counting the people based on motion trajectories, they achieved 94.2% accuracy on the UD-Pedestrian dataset. Another solution for people counting using stereo vision cameras mounted on the ceiling was presented by Van Oosterhout et al. [VOBK11]. The images of the observed area were used to detect human heads in a surveillance scenario. The latter method consists of foreground selection, head detection, and blob separation. As the used data collection did not include attack attempts and the observed area was much wider, a portability evaluation is needed. Hoshino et al. presented a method for people counting and human height measurement using optical flow and stereo vision [HI06]. Their results showed a low average error of 2.5cm in height measurement but had disadvantages related to hair or pose variations. Furthermore, the analyzed scenario focused on moving subjects; it is therefore unknown if this approach will work in the presented scenario. Other approaches are mostly based on the detection of humans from camera side view perspective. The authors of [BBLR05] described an approach for pedestrian detection in

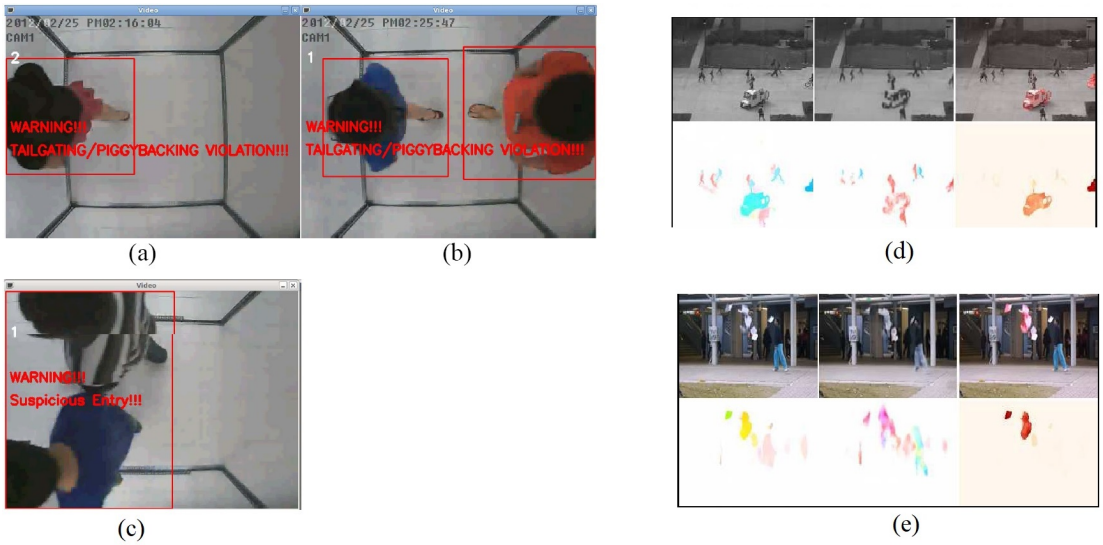


Figure 2.5. – The images a-c show various tailgating/piggybacking detection warnings: (a) warning when number of people is more than one; (b) warning when there are two objects detected in the surveillance area at one time; (c) suspicious entry warning when an object with big contour is detected, e.g. two persons walking side by side passing through the surveillance area [CYS12] - The images d and e show detected anomalies on surveillance camera video data d) The appearance of a truck and a bicycle. (Ped2) (d) A man is tossing papers [NM19].

stereo infra-red images. It is based on the localization and distance estimation of warm areas in the scene. A final validation process is applied based on the head shapes' morphological and thermal characteristics. To identify a human object, a pattern recognition technique is employed where the human head is detected and analyzed for its shape, dimension, and position. Treptow et al. [TCD06] and Ciric et al. [CCNA13] equipped mobile robot platforms with a thermal vision camera for supervisory control and detection of humans. The core of the recognition methods proposed is intelligent segmentation and classification of detected regions of interests in every frame acquired by a thermal vision camera.

Another multi-model approach is presented by Akati and Conrad [AC21]. They propose a three-step anti-tailgating solution that combines face detection, palm recognition, and motion sensors to eliminate the loopholes of existing technologies. The results from experimentation indicated that the face detection tool could detect two faces present. The motion sensors were shown to be efficient in performing people counting and detection, to eliminate tailgating and discrepancies in the number of entries against the number of authorized personnel. They concluded that their system will be an improved and more effective anti-tailgating solution compared to other single-sensor approaches. When analyzing sequences of 2D images, change detection using a Gaussian mixture model was used to detect and count contours [CYS12]. M. Rauter introduced a motion-based head-shoulder detector to detect intrusion [Rau13]. Many approaches only use video frames instead of complex 3D or even thermal data. However, an important factor here is the frame rate, which must ensure that the moment of the attempted attack is recorded with as many images as possible [Tuo19]. The topic of computing time and usability in single-board computers is therefore an important requirement. Optical flow is a strong feature descriptor as it extracts motion and direction besides its position. It is often used in such application to detect motion. By comparison, change detection uses an adaptive background model. This method is more computationally efficient

and allows real-time calculation as we will see later in Chapter 4.6. Another advantage over optical flow is that the data can be augmented more easily because the original feature space is preserved.

A drawback of all these methods is the camera angle, which allows people to hide on the floor or between the legs of a permitted person. Sensors mounted in the floor can be of great help here but are limited to the fact that all feet must be on the floor. Our presented study in Section 4.5 uses capacitive sensors on the ground to detect and count feet [SDF*18] on the floor. It is built upon a floor-based indoor positioning system in grid layout [BHW11]. Its active capacitive measuring system [BWKF15] is efficient for remote sensing and can reliably recognize a foot in up to 10cm height (see Figure 2.6). An obvious limitation is when two people are standing with only one foot each on the sensor surface for which a top-mounted camera will be good match for this application setup.

Figure 2.6. – Schematic of floor based sensing grid output [SDF*18].

Precise localization within a closed building or apartment is of vital importance. It offers a wide range of new opportunities regarding smart home applications and efficient energy control. Floor-based indoor localization systems have the advantage of unobtrusive detection or sensing of the surrounding. They mostly apply pressure-based measurement principles or capacitive sensing technologies. Feng et al. [FYGW15] offered a proof-of-concept study applying fiber optics for localization. Their grid displacements of optical fibers are distributed on the floor. Due to the pressure applied to the fibers through step motions, the signal throughput changes and thus allows the system to localize a person within the measuring area. However, the disadvantage of such a system is its maintenance. Exposed to external forces, the sensors can easily become damaged. To overcome the limitations of a pressure-based system, various floor-based capacitive sensing has been announced with promising results in recent years. SenseFloor [SL08], introduced by Steinhage et al., works with modular capacitive measurement units to detect the presence and location of a person walking within the sensing environment. This system unobtrusively senses the presence of individuals. Another floor-based indoor positioning system using a grid layout instead of modular setups, called Capfloor [BHW11], was introduced by Braun et al. in the year 2011. The advantage of this system is its easy maintenance, since a malfunctioning sensor can easily be replaced instead of a whole floor module. The resulting system is more efficient in terms of power consumption while maintaining the precise localization ability [BWKF15]. Overall, active capacitive measuring systems are more efficient for remote sensing and therefore more robust against pressure. Floor-based systems encouraged us to develop a system based on active capacitive sensing embedded in the floor to solve the problem of recognizing tailgating issues. However, most research on capacitive indoor localization systems offers only low resolution due to the large size of the measuring electrode used, which is inadequate for our targeted application. TileTrack as introduced by Valtonen et al. in [VMV09] can locate a standing human with at least 15 cm accuracy by using 9 separately controllable 60 cm x 60 cm tiles placed in a 3x3m square area. However, TileTrack as well as Capfloor are not able to locate people standing close to each other, which is needed in our use case. A far better resolution

than ordinary capacitive indoor localization systems is therefore needed. To overcome this limitation, we built a system using better resolution cells. The classification method used in their methods is furthermore hardly transferable to our method. As single values are analyzed in Capfloor and TileTrack, this method seems costly with higher resolution.

2.3. Textile Defect Recognition

The impact of artificial intelligence on the industrial sector has exceeded expectations, with numerous researchers and engineers working to accelerate the development of industrial intelligence. Industry 4.0 was introduced by Germany in 2010 and has since been widely adopted in the European Union, followed by similar plans in the United States and China. Industrial artificial intelligence combines mechanical, data science, network, communication, and information security knowledge to improve manufacturing efficiency and security through the use of AI algorithms. Textile manufacturing, which involves complex and ordered processes, must find its way to adapt to Industry 4.0. Fabric defects are a significant problem affecting textile quality, with defects caused by many factors such as material quality, mechanical factors, dye type, yarn size, and human factors. Early detection of fabric defects is crucial to reduce enterprise loss, and automatic fabric defect detection is necessary to reduce costs and increase productivity. Industry 4.0 represents the direction in which the manufacturing industry is headed, and industrial artificial intelligence is a multidisciplinary field that employs AI algorithms to solve industrial problems and enhance manufacturing efficiency and security. As the textile manufacturing industry adjusts to Industry 4.0, it must establish its own adaptation approach. Textile manufacturing is a complex and extensive industry, with a production process that includes spinning, weaving, dyeing, printing, finishing, and garment manufacturing, and the stability and quality of the textile fabric produced by the entire production line are critical to any enterprise. Many factors influence the final product, including material quality, mechanical factors, dye type, yarn size, and human factors, and fabric defects, which primarily result from process problems and machine malfunctions, will impact the final product's quality, resulting in the waste of resources. Early detection of fabric defects is critical in reducing enterprise losses, and effective fabric defect detection is critical for modern fabric manufacturers to control costs and increase product value and core competitiveness. Automatic fabric defect detection is critical in modern textile manufacturing to ensure fabric quality. In addition to traditional and classical methods, deep learning algorithms in defect detection are discussed and compared. The algorithm's deployment is also crucial to the system's accuracy and efficiency. Web inspection is a common application of automatized textile defect inspection. It is mostly performed on spread fabrics and is carried out during their manufacturing process. The most recent work has focused on defect detection and classification. Mishra et al. [Mis15] distinguish woven, knitted, and dyeing/finishing defects which occur during spooning or weaving. Overall, there are about 70 different types of defects in the literature [Cou00]. Some examples of typical textile defects are shown in Figure 2.8. Textiles can be categorized generally into uniform and different kinds of textured materials (uniform, random, or patterned) [Kum08]. For the detection of defects on uniform textured fabrics, three defect detection techniques exist: statistical, spectral, and model-based [NPY11]. But they could also be divided into traditional and learning-based algorithms . Most of the traditional algorithms are based on feature engineering with prior knowledge, covering statistical, structural, spectral, and model-based methods (learning-based algorithms can be further divided into classical machine learning algorithms and deep learning algorithms (See Figure 6.1).

Liu presented a multi-stage unsupervised learning approach [LZF*19] using extreme learning machine classifier training and Bayesian probability fusion. This approach showed good performance on the popular TILDA dataset but suffers, like many methods, from a relatively limited variety of textile defects. Other approaches based on CNNs suffer mostly from the limited amount of data in general, which results in less accuracy [AG19].

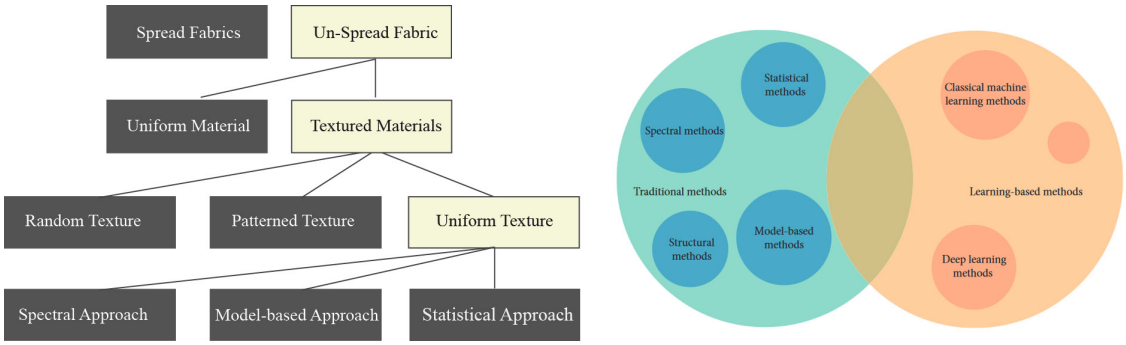


Figure 2.7. – (left) Categorization of web inspection based on the nature of the surface [SKH16](right) Different types of defect detection algorithms [LLL*21].

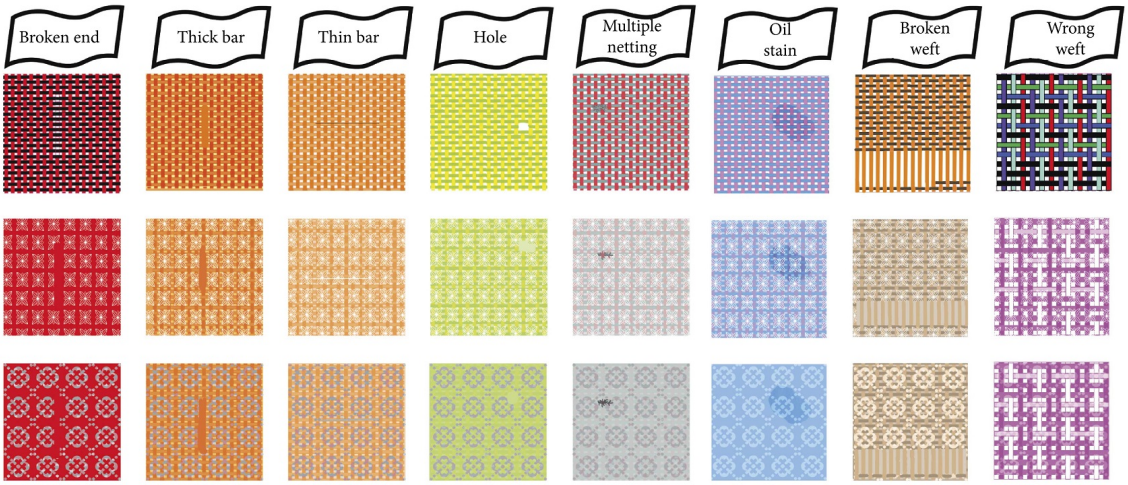


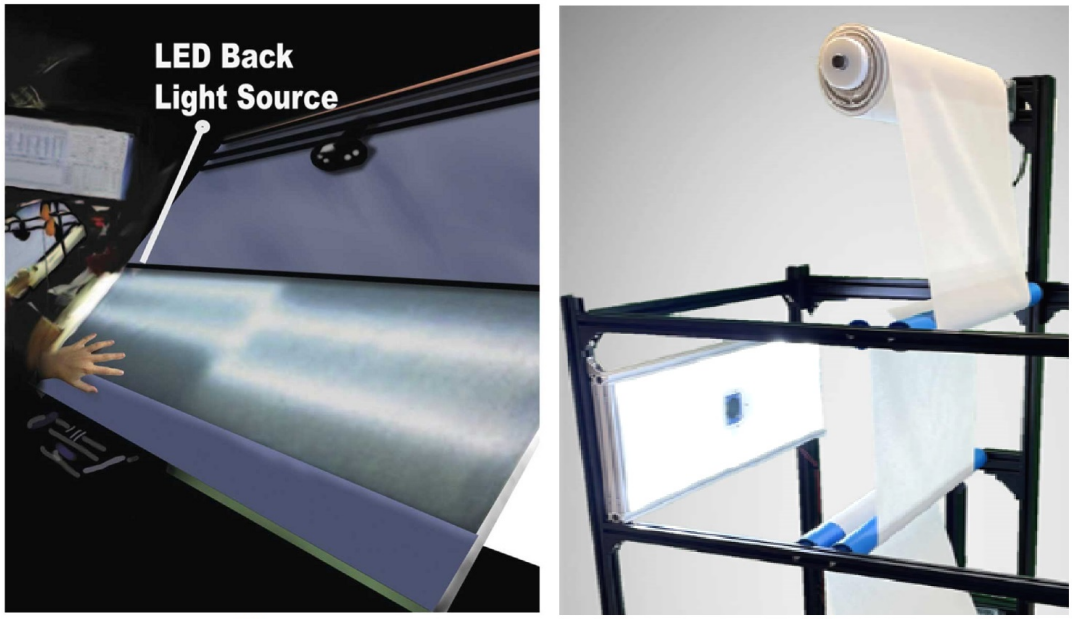
Figure 2.8. – Some common textile defects caused by different reasons of the classes major, minor and critical [RZR*20].

In general, Oni concludes that existing fabric defect detection classification approaches are still prone to a lack of consistency in the image dataset and a challenge for high quality images [OOA*18]. Fabric texture transformation, regular and irregular patterns, a homogeneous spatial relationship between intensity values, and repeated distance units as well as effects caused by the deformable nature of garments are some of the challenges in this field. Several categories of patterned fabric defects are also often not obvious against the texture background. Zhao et al. tackled that problem by presenting a supervised learning strategy [ZHH*20] based on a visual long-short-term memory-based (VLSTM) integrated CNN model. Experimental results on three fabric defect datasets have shown that the proposed model provides competitive results to the current state-of-the-art methods on 2D fabric defect classification. Jun et al. proposed a framework utilizing machine learning to automatically detect fabric defects. The framework has two main steps. First, the original image is cropped using a fixed-size square slider with a certain step and regularity, and then an improved histogram equalization technique is used to enhance each cropped image. Next, the Inception-V1 model is used to predict the existence of defects in the local area. Finally, LeNet-5 is used as a voting model to recognize the type of defect present in the fabric. The proposed framework includes local defect prediction and global defect recognition as its two main steps and achieves state-of-the-art results [JWZ*21]. Another deep-learning-based approach is proposed by enhancing the

YOLOv5 object detection algorithm. A teacher-student architecture is utilized to overcome the lack of fabric defect images. Specifically, a deep teacher network can accurately recognize fabric defects, and the shallow student network can perform the same task in real-time with minimal performance degradation after information distillation. Additionally, multitask learning is introduced to detect both ubiquitous and specific defects simultaneously. To improve the recognition performance, focal loss function and central constraints are incorporated. The results show that it outperforms other methods and exhibits exceptional defect detection ability in the collected textile images evaluated on the public available TILDA dataset [JN21]. Defect detection on (un-spread) deformable textiles in voluminous shape is a relatively new field and adds even more difficulty to the task. Our work focuses on the inspection of textured material which has an almost homogeneous color and uses a combination of a statistical and model-based approaches.

Neural networks (NN), AdaBoost [BF15] and support vector machines (SVM) [MBR04, ZCM18] are notable machine learning techniques that were used in a number of articles in this field. Some approaches on flat, and spread-out 2D surfaces achieve success rates higher than 90 % in fabric defect detection [NPY11, RBAF15, TAS19]. By comparison, humans achieve detection rates of only 60-75 % [Sch93]. In the case of un-spread (in- homogeneous) textile classification, we proposed a system for classification of textile fibers using LBP-features and local-interest points in our recent work [SKH16]. The process was evaluated using preselected image patches in order to reduce computational costs in the textile classification using SVM and AdaBoost. The most time-consuming processes when using these supervised machine learning methods are the acquisition and labeling processes. Thousands of patches must be acquired and labeled manually. In contrast, our novel method reduces the required effort to be spent in labeling data and combines it with convolutional neural network classification. Two of the most challenging problems in voluminous fabric classification are ambient occlusion and folds. These effects are caused by the shape of the textile and the influence of illumination. In a recent work [SBK16]], a normalization method was introduced that reduces these effects, paid by a loss of information. This method, based on stereo-imaging is used in our work for preprocessing of the acquired images. Using disparity information as an additional dimension into a deep learning architecture was investigated in our most recent work [SPKK18]. We have found that the amount of data used to train a network eliminates the need for additional depth information for this application. If sufficient data is available, deep neural networks can learn different deformations such as folds and shadows implicitly and thus become resistant to them. Some approaches utilized artificial neural networks (ANN). Kang and Kim [KK02] involved a trained artificial neural network for color grading raw cotton. The images are captured by a color CCD camera. Furthermore, they acquired color parameters, checked connectivity, and evaluated trash particles for their content, size, size distribution, and spatial density. Their conclusion is that the application of an artificial neural network for color grading is highly valid. She et al. [FSK02] classified two kinds of animal fibers objectively between merino and mohair. In their approach, they developed a system that uses an artificial neural network and image processing for this classification. The image processing extracted the required features for the ANN. Their conclusion is that the classification accuracy of ANN will be improved by developing more powerful learning strategies. Kuo and Lee [KL03] developed a system to distinguish fabric defects like holes, oil stains, warp-lacking, and weft-lacking. For that reason, they used a back-propagation neural network which gets an image as input. They successfully determined non-linear properties and improved recognition. Srikaew et al. [ASK11] presented a hybrid application of the Gabor filter and two-dimensional principal component analysis (2DPCA) for automatic defect detection of texture fabric images. With a genetic algorithm (GA) based on the non-defect fabric images, they achieved the optimal network parameters. With their experiments, they concluded that the applied Gabor filters efficiently provide a straightforward and effective method for defect detection using a small number of training images but still can generally handle fabric images with complex textile pattern backgrounds. Another approach from Sun and Zhou [SZ11] used a threshold segmentation method to identify if there were any defects in the fabric, adopting an image-feature-based approach to recognize oil stains and holes and using training-based techniques

to detect broken ends and missing picks. They segmented and filtered the defect image, extracted features of the fabric defect, and the classification was based on the local features and training. For automated visual inspection, Ngan et al. [HNN05] used the wavelet transform. With direct thresholding (DT) and a so-called golden image subtraction method (GIS), they effectively segmented out the defective regions on patterned fabric. They also present a comparison with other methods.



(a) (b)

Figure 2.9. – Fabric defect detection using (a) backlight [RZR*20] and (b) direct light.

A lot of research effort was spent on fabric web inspection of spread fabrics, carried out during the manufacturing process. Most of them focused on defect detection and classification. Figure 2.9 shows two automatic fabric defect systems. In Figure (a), a backlight is used in order to highlight defects in the fabric; in Figure 2.9 (b), direct light is used. Both light settings have advantages and disadvantages and are therefore often combined. In both cases, humans or area scan cameras, based on CMOS or CCD, can be used. Textiles themselves can be categorized into uniform and different kinds of textured materials (uniform, random, or patterned) [Kum08]. Mishra et al. [Mis15] distinguishes woven, knitted and dyeing/finishing defects which occur during spooning or weaving. For detection of defects on uniform textured fabrics, three defect-detection techniques exit: statistical, spectral and model-based [NPY11]. Neural Networks (NN) AdaBoost [BF15] and support vector machines (SVM) [MBR04] are notable machine learning techniques that were used in a number of articles in that field. Supervised learning strategies achieved a good performance using a counter-propagation NN, trained by a resilient back-propagation algorithm [IAMA06]. As several NN suffer from high sensibility and require a lot of training image samples, we decided to use other machine learning techniques. In the work of Sun and Zhou [SZ11], a threshold segmentation method was used to recognize defects in the fabric. An adoption of image-based features recognized oil stains and holes. The classification of the fabric defects was based on local features and training. A problem which occurs in voluminous fabric classification is shadow, which impairs the classification quality. A normalization technique using color information to classify the physical nature of edges in images was presented in [GS03]. A parameter-free edge classifier for shadow geometry can be calculated but hardly differentiates be-

tween stains and shadows. Kumar assumed in his work [Kum08] that the image quality is highly affected by the type and level of illumination in industrial textile inspection. A comprehensive study of various lighting schemes for automated visual inspection was carried out in [Bat85]. Their results on elimination of shadow and glare effects using backlighting influenced the setting of the capturing environment in this work.

2.4. Summary

This chapter discussed the background information that is needed to understand the research topic of this work and its proposed solutions. Section 2.1 has reviewed (1) the state of the art in the area of biometrics and here specifically in face recognition, presentation attack detection, and handwriting writer recognition. In face recognition, various works seem to have used images in other spectra, but mainly in the infra-red range. It showed potential to improve the performance and reliability of face recognition and PAD as they provide more information which is not detectable in the visual band. Therefore, if this information is complimentary and can be robustly read out, it could increase the accuracy of biometric face recognition systems. It is therefore interesting to investigate the special properties of the ultraviolet range. Since the properties in the UV range are not visible to humans, they could be particularly suitable for use in live recognition and PAD. For this reason, we have created a new database containing images of people of different skin types in the ultraviolet spectrum. In addition, various common presentation attacks, such as masks or 3D printed faces, were recorded in the same spectrum.

Finally, we will present methods that use this data to improve the described drawbacks in the applicability of biometrics. In the case of identifying people by their handwriting, a lot has happened since the advent of deep learning. Even a small amount of text is sufficient to recognize and distinguish biometric features in handwriting. However, filled-in forms represent a special use case. Here, certain letters and characters must be filled in at fixed positions. However, there are also hardly any coherent longer text areas that allow these algorithms to be used in any case. For this reason, it is necessary to investigate whether conventional low-level methods can be used for the evaluation of this data in order to ensure its applicability. In Section 2.2 the current state of the art in the field of autonomous entrance portals was described. Tailgating attempts, in particular, still represent an insurmountable obstacle to limiting access to areas to a certain group of people. Existing hardware and software solutions are not yet able to detect these attempts. Approaches from the field of video analytics for counting and recognizing persons on surveillance videos seem promising. The question is, however, whether a single technology alone can detect these attempts? For this reason, we have created a new database that allows to explore the applicability of different sensor technologies. We have also developed algorithms that use these data specifically for this use case and put them against each other. The third Section (see Section 2.3) has given an overview of the state of the art in textile defect detection. A number of algorithms have been presented that have shown good results on datasets with spread out textiles in this area. Unfortunately, experiments have shown that these approaches cannot be applied to voluminous cloths. The generalization of the existing algorithms to apply defect detection to cloths that are in the form of piles has not been sufficiently considered so far. Furthermore, the question arises whether the setup used for spread materials is also the right setup for this application. Therefore, both new approaches to improve the transferability of the algorithms to non-spread cloth are undertaken, but also other forms of data acquisition are explored.

3. Novel Methods within Biometrics.

In the last chapter, the current state of the art on the three application areas "biometry", "autonomous entrance control", and "textile defect detection" was presented. In this chapter, we will now specifically address three technical innovations that are expected to improve the applicability of biometrics in industry. For this reason, the goal of this chapter is to answer the following research questions:

Research Question 1: Is melamine face pigmentation a biometric modality and can it improve conventional facial recognition algorithms?

Research Question 2: Are there novel feature types or modalities that can be used to increase PAD performance?

In order to answer these questions, we worked exemplarily on three topics and examined them under the focus of practical applicability. The first two topics are in the broadest sense about face recognition, which is one of the oldest modalities in biometrics. It has been used in forensics since the 19th century and is now commonly used on smartphones. Nevertheless, especially for this reason, the use of many millions of people has created a need for even better recognition methods. Melanin face pigmentation (MFP), or colloquially "freckles", are skin phenomena that appear visibly or invisibly on some people. Whether the targeted evaluation of these characteristics can be used to improve the performance of a biometric system is answered in Section 3.1. Can MFP perhaps even fulfill the requirements for a modality of its own? Can melanin face pigmentation (MFP) contribute to an improvement in face recognition?

As promising and effective biometric solutions have proven to be, they have become nonetheless subject to fraudulent attacks. Therefore, the vulnerability of such systems against fake biometric characteristics is a growing concern. The so-called presentation attacks can be expressed in terms of, but not limited to, someone posing as another individual or hiding their identity [TLLJ10]. Subsequently, that led to the development of “presentation attack detection” (PAD) or “anti-spoofing” techniques. A prerequisite to a good and reliable PAD application is most notably the ability to perform well with different kinds of attacks and scenarios under diverse conditions. Whether melanin face pigmentation is suitable as a PAD to detect attacks on biometric systems will be described in Section 3.2.

However, PADs are not only a problem with face recognition but also with other modalities, such as handwriting. Recognition of individuals using handwritten text or signatures is an established topic in biometrics because of the social and legal acceptance and the widespread use of documents and forms filled by hand [HSBAF22, RNR19]. Handwriting on documents and especially forms with different text elements can often be distinguished. But what happens if a person fills out a form multiple times to indicate that different people have rated a product? If this occurs within the framework of a biometric system or a biometric examination, this is called double enrollment. In Section 3.3 I show whether it is possible to recognize similarities in handwriting using forms and to compare them in the form of a double enrollment check.

3.1. Melanin Face Pigmentation (MFP)

Facial recognition in the visible spectrum is a widely used application but it is also still a major field of research. Melanin face pigmentation (MFP) can be seen as a new modality to be used to extend classical face biometrics. Melanin pigmentation are sun-damaged cells that occur as revealed and/or unrevealed pattern on human skin. Most MFP can be found in the faces of some people when using ultraviolet (UV) imaging. To proof the relevance of this feature for biometrics, we present a novel image dataset of 91 multiethnic subjects in both, the visible and the UV spectrum. We show a method to extract the MFP features from the UV images, using the well known SURF features and compare it with other techniques. In order to proof its benefits, we use weighted score-level fusion and evaluate the performance in an one against all comparison. As a result we observed a significant amplification of performance where traditional face recognition in the visible spectrum is extended with MFP from UV images. We conclude with a future perspective about the use of these features for future research and discuss observed issues and limitations.

We present a solution that tackles PAD, by using multi-modal biometrics in the ultra-violet (UV) spectrum by analyzing Melanin Face Pigmentation (MFP). As discovered in a recent paper [SSG*18], MFP can be seen as additional modality, of which most can only been seen by a sensor, sensible in UV wavelength. In this paper we analyze if these captures of the human skin are useful for PAD by presenting a novel method on this application. The first method detects the MFP in the images using ORB keypoints and tries to identify attacks using their number and distribution. In the second method, we examine whether the corresponding brightness can be used as a feature between the images. Both methods use two captures at the same time, one in the UV spectrum and another made in the visual spectrum (VIS). To confirm our assumption that PAD works by using images in the UV spectrum, we present a novel database of presentation attacks that includes images in UV and VIS spectrum (see Section 3.1.1). This database contains images of 2D prints on paper, 3D prints and masks of different material. These images are evaluated together with a recently published database of 91 real subjects captured over a period of 6 month showing different expressions and poses. The methodology of our verification methods is presented in Section 3.1.2. There, we describe the images descriptors and fusion methodology that we used in our methodology. Our results in Section 3.1.6 show if UV face imaging and/or MFP, provide valuable distinct information for face PAD. We conclude with a future perspective about the use of these properties for future research and highlight observed issues and limitations in Section 6.1.1.

3.1.1. Database

As we could not find a public dataset with face images in the UV spectrum, we decided to collect a novel dataset. We captured simultaneously, images in the UV, as well as in the visible spectrum under conditions, as we would expect them in an controlled scenario, such as border control. We captured people of different skin types, age and gender and let them classify their skin according to the Fitzpatrick Scale [Fit88].

3.1.1.1. Recording Setup

We used two synchronized cameras, attached side by side, in order to keep the divergence in perspective small. The test participants were positioned at a set distance of 1.5m away from the cameras, illuminated with light in visible and UV spectrum. The position of the used lights was chosen in a way that shades are similar in both captures. In order to avoid interferences, UV/IR and VIS filters were used respectively, allowing only the transmission of the intended wavelength. The used setup allowed a simultaneous capturing process, without changing filters and lights.

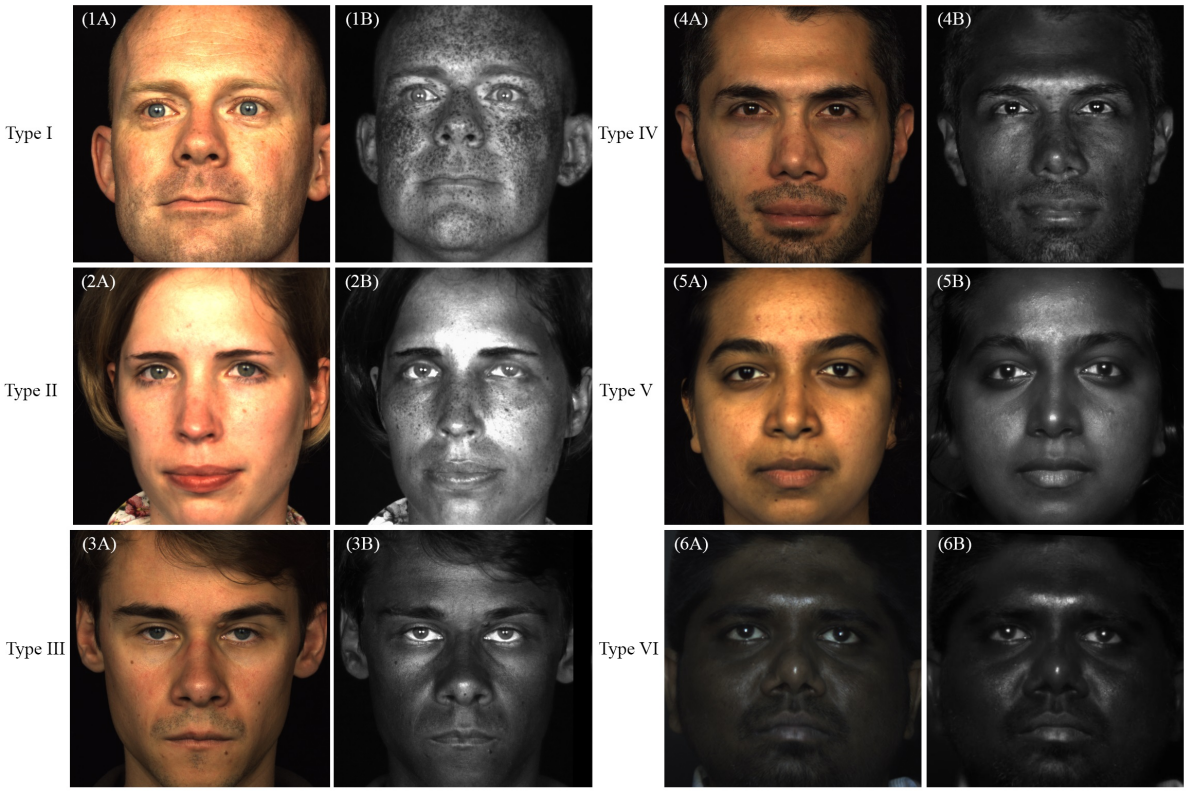


Figure 3.1. – Face samples of different skin-types (according to Fitzpatrick Scale) in VLA (column A) and UV (column B).

For the UV capturing, a Baumer VCXG-13M camera with a 1/2" CMOS sensor, 1280x1024 pixel resolution and a sensitivity starting at 300nm was used. In order to receive best response of UV-MFP, we filtered the visible light in the captures by using a UV/IR bandpass filter (Schott UG11) which has its peak transmission at 300nm and blocks light with wavelength over 400nm. Furthermore a quartz lens was used to increase the light transmission in that wavelength. For illumination we used two 36W UV-A LPS lamps with a bandwidth between 315nm and 400nm positioned in front left and front right to the subject. In Figure 3.2 the sensitivity of the camera sensor, the transmission of the used filter and the emission of the used illumination in UV spectrum is shown. Due to different sensor size, focal length and arrangement, the captures show little perspective distortion and a slightly different angle of view.

In the capturing of the visible light images a AV Prosilica GT 3400 with a 3384x2704 pixel sized CCD sensor and a regular UV/IR cutoff filter was used. Two diffused studio lights with diffused 8x70W light bulbs (each 3570Lm) were used.

3.1.1.2. Difference to other recording setups

In comparison to the data shown in an earlier study of Narang et al. [NBH15], we achieved a better resonance of UV-MFP in our images. A possible reason for that is the use of the UV bandpass filter. As camera sensors

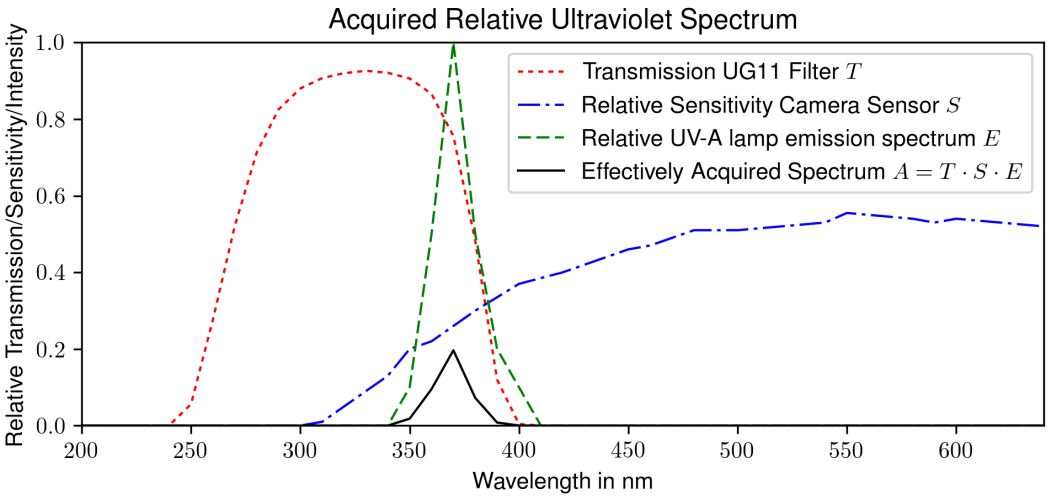


Figure 3.2. – Transmission of UG11 Filter, Sensitivity of Camera Sensor and Emission of UV-A light in different wavelengths.

have a significantly higher sensitivity above 400nm wavelength and most lamps also emit light in that spectrum, information in the UV spectrum become less dominant in the image. It is also important to set the exposure time individually, so that UV-MFP get clearly visible. Besides that, the approach on creating gallery and probe images is quite similar.

3.1.1.3. Data Acquisition

We captured 924 portrait images of 91 people in UV and VIS, with a resolution of 640x512px and 1021x944px respectively. All images were aligned and resized to 600x600px. People of different ethnicities participated in the capturing process, in order to cover all skin types of the Fitzpatrick scale [Fit88] (see Figure 3.1). The age of the participants was between 15 and 62 with a gender distribution of 27%/73% female/male. We captured the data in three sessions, in mid-March, end-July and mid of September 2017 in Germany. Some of the test participants used make-up and/or suncream (see Figure 3.3) or acquired a tan between the different recording sessions. In each session images with regular expression, with a smile expression and an expression chosen by the participant was taken. In order to explore the effect of different recording angles, we captured images from left, right, top, and bottom directions with an angle of approximately 45°in the second session.

As presentation-attack-detection using UV cameras might be a topic of future research, we also captured images of common spoofing attacks. We used latex and silicone masks as well as printouts, in a variety of different paper and printer types of some of the participants. The face of one test participant was printed in a 3D printer with white PLA filament, the face of another participant was molded in a silicon mask.

3.1.1.4. Image Preparation

All full-resolution VIS and UV images were aligned and cropped to the face region by using face detection and alignment by Zhang et al. [ZZLQ16] which was significantly more capable of aligning the UV images then using

Figure 3.3. – Effects of suncream on the UV image.

Eigenfaces [TP91]. We were still forced to manually align 25 UV-images to guarantee their meaningful inclusion into the dataset. No VIS images were aligned manually.

3.1.2. Evaluation Method

To prove that there is significant information in the ultraviolet spectrum for biometrics, we tested our data with local features. The importance of local patterns was described from Mikolajczyk et al. [MS05] and Penev et al. [PA96] in great detail. We expected to find MFP in high frequency features with a pixel size between 3 and 20 pixels (px).

Evaluating current face recognition performance is done mostly on big datasets that are widely available for the visible light spectrum, e.g. the Labeled Faces in the Wild database [HRBLM07] or the YouTube Faces [WHM11] dataset. A comparable dataset was not found for the ultraviolet light spectrum which led the authors to the creation of a dataset as described in Section 3.1.1. Evaluation was aimed at providing evidence for valuable information in the ultraviolet spectrum because the size of the dataset would yield nearly perfect results in a verification scenario with modern approaches in face recognition.

3.1.3. Face and MFP Descriptors

We analyze the available front images, with and without expressions from our database in a one versus all (OVA) experiment. The chosen algorithms types therefore differ in computational speed, rotation variance and size tolerance. Three different feature types are used in our experiments: LBP (Local Binary Patterns) [OPM02], SURF (Speed Up Robust Features) [BTVG06] and CSLBP (Center Symmetric LBP) [HPS09]. CSLBP and LBP are chosen because they are well known and relevant image descriptors in the filed of face recognition [AHP06]. In contrast to that, has SURF shown its performance in different applications like: locating and recognizing objects, people or faces. The SURF keypoint detector is rotation invariant and uses an approximation of the

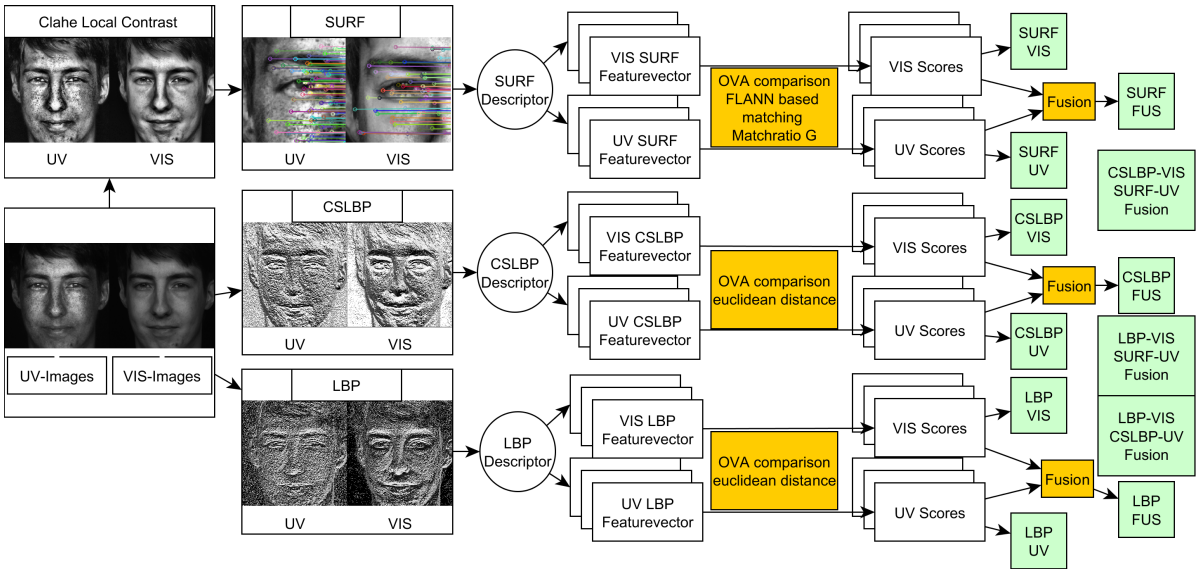


Figure 3.4. – Diagram showing the used pre-processing, image descriptors, comparison metrics and fusion experiments.

determinant Hessian blob detector to find relevant points of interest. As most MFP in human faces occurs as small blobs like points, we expect a high resonance with that method.

To make the performance comparable, the descriptors are built by using the best case parameters for each experiment according to their original papers [OPH96, HPS06, BTVG06]. In the case of SURF, local contrast enhancement was performed (see Figure 3.4), using a local histogram mapping function [JSR94]. The ideal combination of parameters was compiled throughout the progress of the work.

3.1.3.1. LBP Feature Vector

LBP features are calculated for results in 256 values for each pixel. The feature matrix is divided into 16px $\times$ 16px matrices. For each matrix a histogram with 256 values is calculated and concatenated into one descriptor. This feature vector of 1406 matrices $\times$ 256 dimensions results in 359936 values. The euclidean distance between two vectors is used as the comparison score.

3.1.3.2. CSLBP Feature Vector

The features created by CSLBP have only 16 dimensions and are built as described by Heikkilä et al. [HPS06]. In the publication, local and point symmetric binary relations are used to construct a feature with 16 bits for each point. Every bit describes the relation of two point symmetric neighbors of a target point. The best performing parameters in the publication consist of the radius $R = 2$, the neighbors $N = 8$ and the threshold $T = 0.01$. The created matrix is divided into 60px $\times$ 60px matrices which produces 100 histograms concatenated to a feature vector of 1600 features. To equalize the descriptive force of every histogram bin, the histogram values v_i are normalized to unit length, cropped to the values $0.0 < v_i < 0.2$ and normalized again to unit length. This is done in order to reduce the influence of large gradient magnitudes. Once again the euclidean distance between two vectors is used as the comparison score.

3.1.3.3. SURF Feature Vector

SURF keypoints are calculated using a Hessian threshold of 400 and eight octaves. Three octave layers are extended and upright matched because a high distinctiveness is needed and the features are expected to be found in similar orientations. In order to find similarities, the FLANN based matching method is used. The quality of matches m in the set of all matches A has to be distilled to the most descriptive points. To achieve this a distance threshold $t(d_{min})$ is used between the matches, where d_{min} is the minimal distance between all matches:

$$s = 0.02$$
$$t(d_{min}) = \begin{cases} (s \cdot d_{min}),\ (s \cdot d_{min}) > s \\ s,\ otherwise \end{cases} \tag{3.1}$$

As an additional refinement of matches in the set, we removed matched points with a spacial distance of 30px or more. This was done to restrict meaningful correspondences between areas of the face image e.g. forehead with forehead and not forehead with mouth. This leads to a Subset of good matches $B \subseteq A$. In order to measure a score the authors used the match ratio $G = |B| / |A|$, where $|B|$ are the number of good matches and $|A|$ are the number of all matches.

3.1.4. Score Level Fusion

Scores and score distributions are in general not directly comparable between different kinds of feature vectors, which makes fusion by sum or normalized sum between e.g. LBP and SURF scores not a viable option for score-level fusion. To use each score type in relation to its performance, normalization and weighting have to be done first. PAN-min-max (Performance Anchored Score Normalization) was used to first normalize each score s out of each OVA score set S_k as described by Damer et al. [DON13]. This shifts the sensitive score range at the threshold of the EER (Equal Error Rate) $T(S_k)$ to a comparable scale and equal position of two sets S_1 and S_2. Thereby k is the respective index for the score sets of each experiment e.g. LBP-VIS or CSLBP-UV. The normalized scores are therefore computed according to the following function:

$$f_{PAN}(s) = \begin{cases} \dfrac{s - min(S_k)}{2 \cdot (T(S_k) - min(S_k))}, if\ s <= T(S_k) \\ 0.5 + \dfrac{s - T(S_k)}{max(S_k) - T(S_k)}, if\ s > T(S_k) \end{cases} \tag{3.2}$$

After normalization each feature type has a specific performance which has to be reflected in the weighting of the scores before summing the fusion. We do this by using the general approach OLDW (Overlap Deviation Weighted) in the publication [DON14b]. In this approach the overlap area between impostor S_k^I and genuine S_k^G scores is used together with an overall performance measure, the EER, to construct the OLD_k measure for each set in the experiments. In this case the EER-Threshold $T = 0.5$ because the values are normalized according to the EER in the center of the unit length.

$$OLD_k = \sigma\left(\left\{S_k^I | S \geq T\right\} \cup \left\{S_k^G | S < T\right\}\right) \times EER_k \tag{3.3}$$

To extract each weighting ω_k for a set in a fusion process we use the equation as shown below:

$$\omega_k = \frac{\frac{1}{OLD_k}}{\sum_{k=1}^{N} \frac{1}{OLD_k}} \tag{3.4}$$

As a result the scores S_k in each set respectively are fused in the following manner:

$$f_{Fusion}(S_k) = \sum_{k=1}^{N} \omega_k \cdot S_k \tag{3.5}$$

3.1.5. Experiments

As described in Section 3.1.2 an one versus all approach is used to test the datasets with respective feature types. Eleven experiments (see Figure 3.4) are conducted and measured by corresponding ROCs, and include an evaluation of each descriptor separately for UV and VIS images and a fused score. Additionally the best performing scores are fused respectively.

3.1.6. Results

We created Receiver-Operating-Characteristic curves in order to show the performance of the different image descriptors (see Figure 3.5 a-c) and fusion approaches (see Figure 3.5 d-f). Every experiment besides the LBP shows an increase in performance when fusing with UV scores. Especially the SURF experiment seems to be triggered higher on the MFP features. This could be due to the nature of SURF being spatially more precise and better aligned on the size of specific local features like MFP. When using SURF on the VIS images, significantly less keypoint will be found compared to the UV images (see Figure 3.4).

In order to make the results comparable, the True Acceptance Rate (TAR) at a False Acceptance Rate (FAR) of 0.01 is shown in Table 3.1 and at a FAR of 0.1 in Table 3.2. TAR is a measure of the probability that an authorized user will correctly accept an access attempt by the biometric authentication system. It is represented as the percentage of times an authorized person instance is correctly recognized by the authentication system. For an authentication system to be more successful, this measure has to be maximized (max=1.0) and can be calculated as in Eq. (3.6).

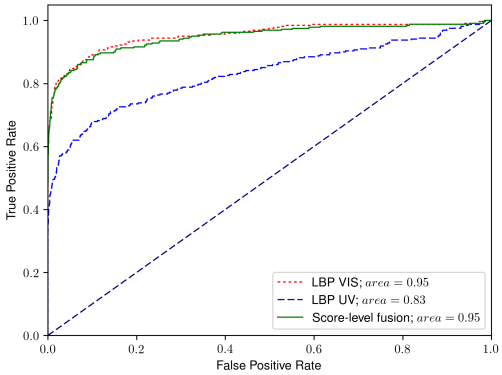
$$TAR = \frac{\text{Number of authorized user instance that is correctly accepted}}{\text{Total number of authorized user instances}} \tag{3.6}$$

The FAR is a measure of the likelihood that an unauthorized user will incorrectly accept an access attempt by the biometric authentication system. It is represented as the percentage of times an unauthorized user instance is incorrectly accepted by the authentication system. For an authentication system to be more successful, this measure has to be minimized and can be calculated as in Eq. (3.15) [DKS*22].

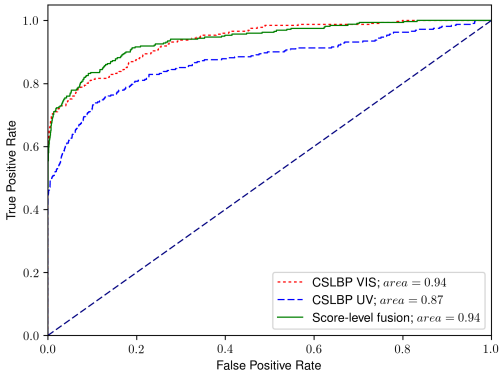
$$FAR = \frac{\text{Number of unauthorized user instance that is incorrectly accepted}}{\text{Total number of unauthorized user instances}} \tag{3.7}$$

In order to make the evaluation comparable, we used fixed TAR of 0.1 and 0.01 and calculated their corresponding TAR.

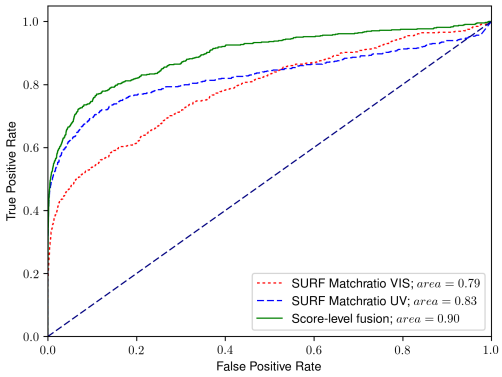
The LBP-VIS/CSLBP-UV experiment also shows an increase in performance at a FAR of 0.01 we see a jump of 0.89 to 0.91. An increase is also visible in the CSLBP-VIS/CSLBP-UV experiment, where the performance



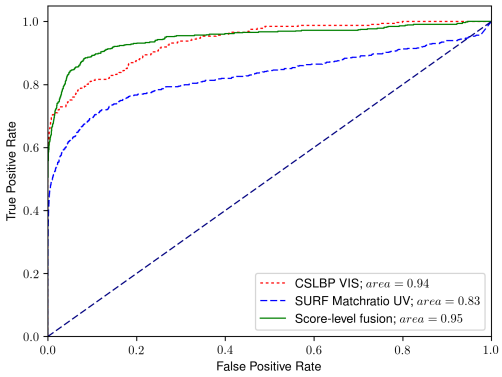
(a) Results using only the LBP descriptor.



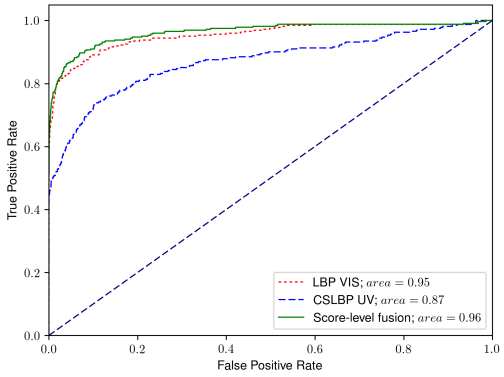
(b) Results using only the CSLBP descriptor.



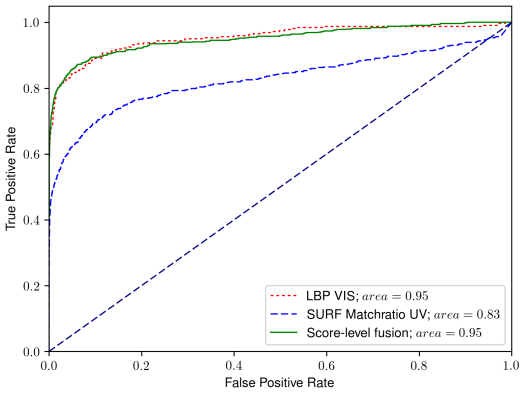
(c) Results using only SURF Keypoints.



(d) Results of fusion btw. SURF-UV and CSLBP-VIS.



(e) Results of fusion btw. LBP-VIS and CSLBP-UV.



(f) Results of fusion btw. SURF-UV and LBP-VIS.

Figure 3.5. – The curves a-c show the results when different image descriptors are getting used. Figures d-f show the results when score-level-fusion is used to combine the different methods.

Table 3.1. – True Acceptance Rate @FAR of 0.01

	CSLBP-VIS	CSLBP-UV	LBP-VIS	LBP-UV	SURF-VIS	SURF-UV
CSLBP-VIS	0.81	0.83	-	-	-	0.89
CSLBP-UV	0.83	0.72	0.91	0.76	-	0.78
LBP-VIS	-	0.91	0.89	0.88	-	0.91
LBP-UV	-	0.76	0.88	0.68	-	-
SURF-VIS	-	-	-	-	0.54	0.75
SURF-UV	0.89	0.78	0.91	-	0.75	0.69

Table 3.2. – True Acceptance Rate @FAR of 0.1

	CSLBP-VIS	CSLBP-UV	LBP-VIS	LBP-UV	SURF-VIS	SURF-UV
CSLBP-VIS	0.94	0.94	-	-	-	0.95
CSLBP-UV	0.94	0.87	0.96	0.89	-	0.88
LBP-VIS	-	0.96	0.95	0.95	-	0.96
LBP-UV	-	0.89	0.95	0.83	-	-
SURF-VIS	-	-	-	-	0.79	0.90
SURF-UV	0.95	0.88	0.96	-	0.90	0.83

rises from 0.81 to 0.83 as seen in Table 3.1. A significant rise is seen in the SURF-VIS/SURF-UV experiment, at a FAR of 0.01 from 0.54 and 0.69 to a fused 0.75. This increase could be indicative for the specificity of SURF-features which respond well with freckle-like round structures. Generally performs the LBP descriptor better on VIS face images than the CSLBP descriptor, while CSLBP descriptor is better at the UV images. We selected the best performing descriptors on the VIS images: LBP-VIS and CSLBP-VIS, in order to fuse them with the best performing UV image descriptors: SURF-UV and CSLBP-UV. As shown in Figure 3.5, is there a significant increase of performance in two of the analyzed approaches, when using fusion, while the fusion of the LBP-VIS and SURF-UV scores remain at LBP level.

During the tests we observed that the skin-type has a high influence on the performance of the MFP feature. While there was not found many keypoints for skin types IV to VI, is this contrarily for skin-type I-III (see Figure 3.1). But because of the small size of our dataset, we can not contribute with a skin-type wise statistics of performance.

3.2. Face Presentation Attack Detection

Face recognition is one of the most commonly used biometric modality. However, the security of the commonly used algorithms is often doubted lately, as they appear vulnerable to the so-called presentation attacks. There are a number of detection methods that are using different light spectra to detect these attacks while this is the first work to explore skin properties using the ultraviolet spectrum. Our multi-sensor approach consists of learning features that appear in the comparison of two images, one in the visible and one in the ultraviolet spectrum. We use brightness and keypoints as features for training, experimenting with different learning strategies. We present the results of our evaluation on our novel Face UV PAD database. The results of our method are evaluated in an leave-one-out comparison, where we achieved an APCER/BPCER of 0%/0.2%. The results obtained indicate that UV images in presentation attack detection include useful information that are not easy to overcome.

3.2.1. Database

The evaluation of the proposed method is carried out on an extended version of a newly created UV-Face database [SSG*18]. The database consists of images collected in the UV, as well as in the VIS spectrum under conditions, as one would expect them in a controlled scenario, such as border control. Compared to the IR bandwidth, one of the first observations when exposing human skin to UV emission is, that skin of different people looks quite different in that spectrum. The Fitzpartick scale[Fit88] groups the skin type into six different categories, according to the reaction of the skin to the sun. Most notable with skin type I, where people show additional MFP in the UV image, that aren't visible for human eyes. The database includes subjects of different age, gender and skin types. We've expanded the database by 127 images of spoofing attacks by using a variety of materials to cover a wide range (see Table 3.3). We evaluated on the following selection of attacks according to reported attacks in media and research.

- A color-bust made of photopolymers, printed on a Stratasys – Connex 3 3D printer, using the polyjet method [BAU15] (see Figure 3.6-1).
- A 3D face bust (17x11cm), printed on a Prusa i3 MK3 3D printer by using PLA (Polylactide) as printing material (see Figure 3.6-2).
- A 3D face bust using silicone rubber on a 3D mold. The mold was created by following this tutorial[Tig19], using alginat for the imprint (see Figure 3.6-5).
- Two unpainted, professional latex masks, as they can be brought on the internet and two painted masks of the same material. We created various variations by combining them with wigs (see Figure 3.6-3 and 4).
- A unpainted mask, made of foam latex.
- Twenty color-printouts, made on a laser-printer. Ten of them were made using normal paper, another ten were made using thicker shiny paper (see Figure 3.6-6).

Table 3.3. – Summary of the Data used in Bona Fide and Attacks.

Bona Fide Images	Identities	UV/VIS
Skintype 1-2	28	238/238
Skintype 3-4	45	521/521
Skintype 5-6	18	165/165
Material	**Identities**	**UV/VIS/Augmented**
3D Print bust (Photopolymer)	1	7/7/14
3D Print bust (PLA)	1	8/8/16
3D Silicone rubber bust	1	7/7/14
3D Latex Mask (painted)	2	28/28/28
3D Latex Mask (unpainted)	2	25/25/25
3D Latex Foam Mask (unpainted)	2	5/5/10
2D Laser Color-Print	10	10/10/10
2D Laser Color-Print (Shiny paper)	10	10/10/10

Each attack is captured by using both cameras in following poses: frontal, 45°view to the left, 45°view to the right, looking up, looking down. Two cameras, attached side by side, are used in order to keep the divergence in perspective small. Test participants, wearing the masks, or the 3D models are positioned at a distance of 1.5m away from the cameras. In order to avoid interferences, UV/IR and VIS filters were used respectively, allowing only the transmission of the intended wavelength. For the UV capturing, a DLP LLC camera with a CMOS sensor, resulting in images of 2592x1944 pixel resolution is used. For illumination we used two 36W UV-A

LPS lamps with a bandwidth between 315nm and 400nm positioned in front left and front right to the subject. The position of the used lights was chosen in a way that shades are similar in both captures. The images in the visible spectrum are captured by using a Nikon D9000 with a APS-C CMOS sensor and a 35mm lens. The UV images are resized by 58% and cropped to 600x600pixel, the VIS images respectively. All images are converted to gray-scale. We augmented the attack database by slightly changing the saturation for every image pair by using linear transformation.

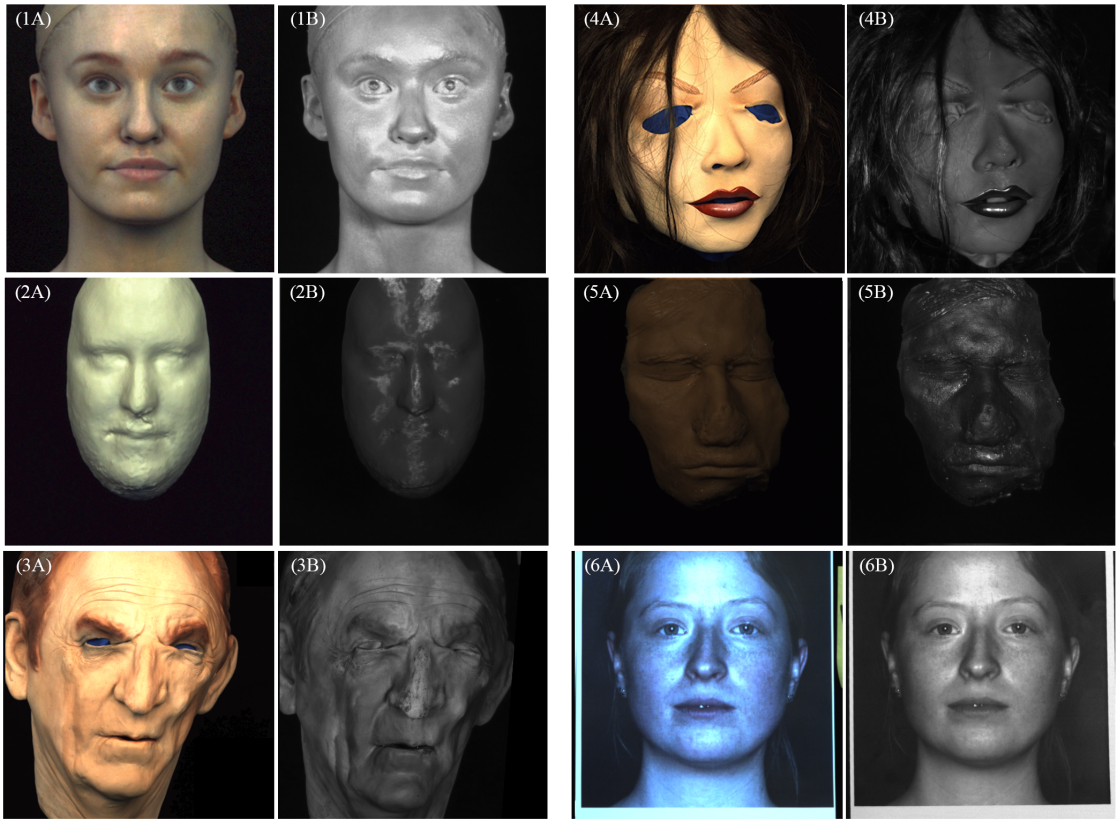


Figure 3.6. – Examples of created spoofing attacks in VIS (A) and UV (B).

3.2.2. Introduced Methods for UV-PAD

As one of the first observations, after capturing the attack images, we found that the brightness of the images differs greatly in UV compared to the VIS. Since both VIS and UV image are taken simultaneously by us, we can rule brightness manipulation by the attacker out. While the brightness of the silicone bust (see Figure 3.6-5B) is relatively low, the 3D color print made of photopolymers 3.6-1B) reflects a lot and is therefore very bright. Of course, it can also be assumed that UV images of non-skin have no MFP, which would be additionally evaluable on the UV images. Another observation is that relatively smooth material, such as latex masks with no notches, have almost no details in the UV spectrum (see Figure 3.6-4B). Furthermore, all latex masks show no reflections that lead to overexposure at all. Comparing that to bona fide images we observed that there is almost no image

that does not show at least a small area like this (very often at the forehead). However, smooth material such as the silicon print, the 2D prints or the PCL 3D print have very strong reflections of this kind. In the case of the 2D printouts, it was even possible only at certain angles to capture images at all, where not the complete face is superimposed by this effect. The main difference between two images is the overexposure in some places, apparently due to the material. However, this effect also occurs in the images of the bona fide group, and is therefore not suitable for a targeted evaluation.

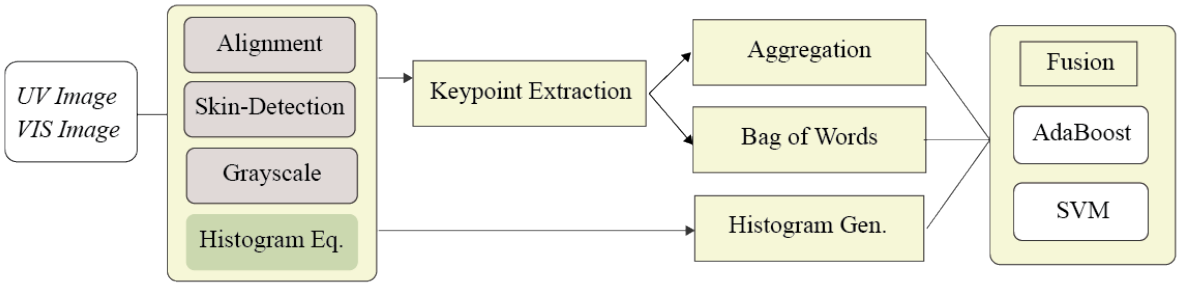


Figure 3.7. – Flow-Diagram of the proposed Methodology.

These observations lead us to evaluate these properties in different ways. If there are no differences between the two images, as would be the case with real skin (MFP), this is an important feature, which can be characteristic of attacks and bona fide. Secondly, there are differences in the ratio of the brightness from VIS to UV, which may differ from those of the skin, they can be seen as spectral signatures. As our database is relatively small, we could not effectively use any deep- or transfer learning approaches. Therefore, we chose conventional features to analyze those characteristics and prove their significance (see Section **??** for a flow diagram of the proposed method). Since both properties only affect the skin, we use the same pre-processing steps for both methods, which is described in the following section. Our method for extracting the different details of both images are explained in Section 3.2.2.2. The brightness differences are presented in Section 3.2.2.3.

3.2.2.1. Pre-processing

Initially, face detection is performed on the full resolution images. After that, VIS and UV images are aligned to the face region by using face alignment by Zhang et al.[ZZLQ16]. We aligned several images manually in order to guarantee their meaningful inclusion into the dataset. Since it is not expected that the eye region will provide valuable information, we remove this region with a mask. Since hair and the mouth region also contain no valuable information, we perform skin detection by using the procedure of Buza et al. [BAO17] and mask-out all non-skin pixel. In a next step, we convert all images to grayscale, in order to reduce the complexity of our small data-set. In our approach, which evaluates the similarity of local features (See Section 3.2.2.2), we also do histogram equalization, which we do not do in the case of brightness analysis (see Section 3.2.2.3).

3.2.2.2. Analysis of Similarity using local Features

As already shown in previous work[SSG*18], MFP features can be extracted effectively by keypoints (KP). We expected to find these properties which are visible in the UV spectrum in high frequency features with a pixel size between 3 and 20 pixels (px). We have therefore selected the ORB (Oriented FAST and Rotated BRIEF) feature detector [RRKB11] to extract this property. The ORB detector is computationally very efficient with similar matching performance to SIFT but less affected by image noise and can be used in real-time. A maximum

of 1000 ORB KP are calculated using the harris score ranking and four points to produce the oriented BRIEF descriptor. Matching is done by using the euclidean distance between two points, one in the UV image, one in the VIS image, assuming that they denote the same feature if the euclidean distance is smaller 10 px. In Figure 3.8 the results of the KP extraction and matching is shown on two images. In the upper images of an attack, with unpainted latex mask, it can be seen that hardly any of them are detected on the surface. It can only be found along the mouth, while in the bona fide image (below) they are recognized throughout all many of them can be matched. The overexposed area on the forehead in the UV image is also clearly visible.

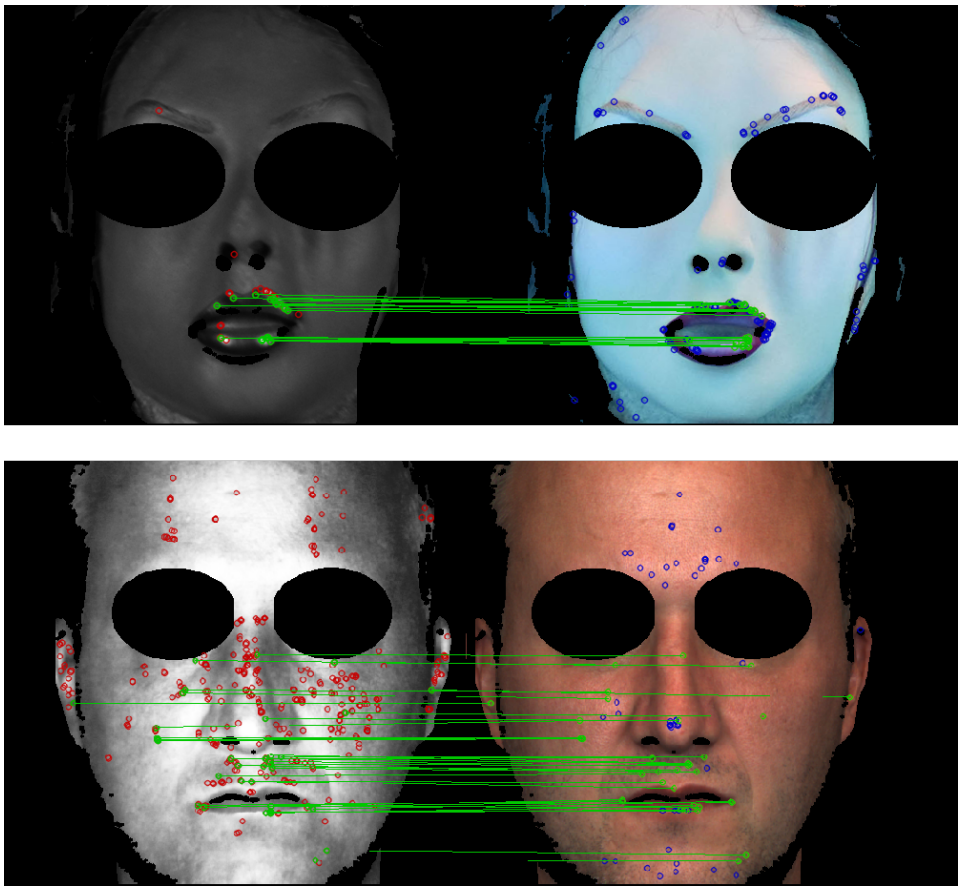


Figure 3.8. – Matched and Unmatched Keypoints in a typical bona fide image of Skintype II (bottom) and an Attack using a Silicone Mask (top).

With the described method we have extracted the keypoints on all image pairs. We can detect the following three main differences between attacks and bona fide: (1) Their number found is smaller for attacks compared to non-attacks (see Figure 3.8A)(2) Bona fide show more KP on the UV image that can not be matched with ones on the VIS image (see Figure 3.8B). (3) In attacks, more unmatched KP can be found on the VIS image than are found on the UV image. The 3D silicone imprint exhibits an extremely high number of unmatched keypoints on both images. These attributes allow us to distinguish both classes in particular, we visualized the number of unmatched KPs in the UV and the VIS image over all classes in Figure 3.9.

We assume that the number of unmatched keypoints between UV and VIS, as well as between VIS and UV contain discriminative information. Therefore, we compose our feature vector as follows: (1) Total KP detected in UV (2) Total KP detected in VIS (3) Matched keypoints (4) Unmatched KP in the UV image and (5) the number of KP in VIS that could not be matched to the UV image.

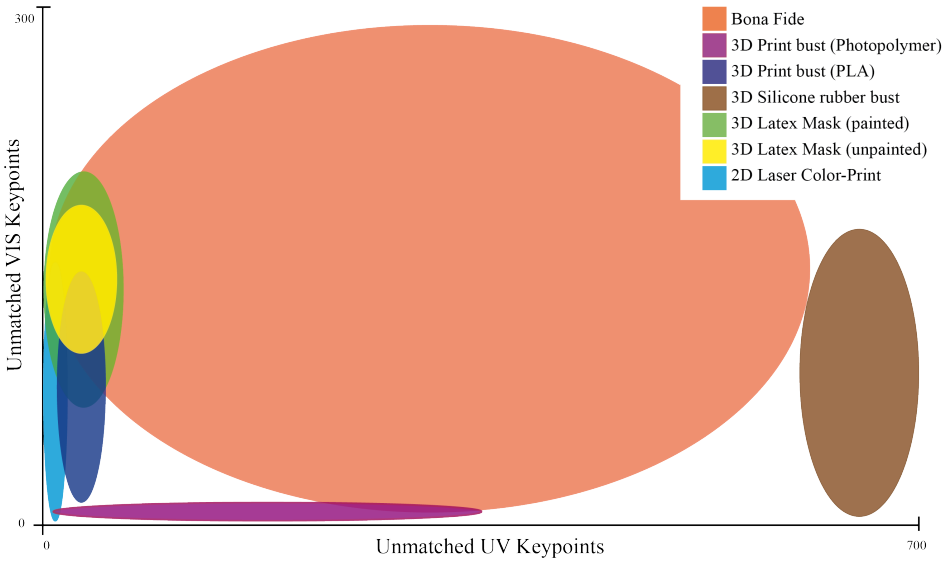


Figure 3.9. – Unmatched ORB Keypoints in the UV and VIS Image.

3.2.2.3. Analysis of Brightness Property

As can be seen in Figure 3.8, attacks reflect differently from bona fide faces when captured with an UV camera. Figure 3.10 depicts the average difference between visual and ultraviolet images of both bona fide faces and attacks presented in a gray-scale histogram. Thus, this method utilizes the distribution of their brightness values in the form of histograms. It aims to discern legitimate images from attacks by comparing the histograms of both the UV and the VIS image of faces. Since the histograms represent the image's brightness distribution, each has a length of 255. By combining both histograms for one face, we create feature vectors containing the amount of pixels that are of each particular brightness for both the UV and VIS image. We experiment with different methods of combining, including adding, subtracting and concatenating the histograms for feature vectors of a length of either 255 or 510.

3.2.3. Experiments and Results

In our approach using keypoints, we experimented with adding and omitting the features written in Section 3.2.2.2. Here, the variant using all five values has proved to be the best. Using the positions of the keypoints, we experimented with different feature vector lengths between 150 and 500. A length of 300 has proven to be optimal. The results are reported in terms of Bonafide (real) Presentation Classification Error Rate (BPCER) and Attack Presentation Classification Error Rate (APCER). BPCER is defined as the rate of the bonafide (real) samples classified as attack samples (see equation 3.8). APCER is the proportion of the attack samples classified

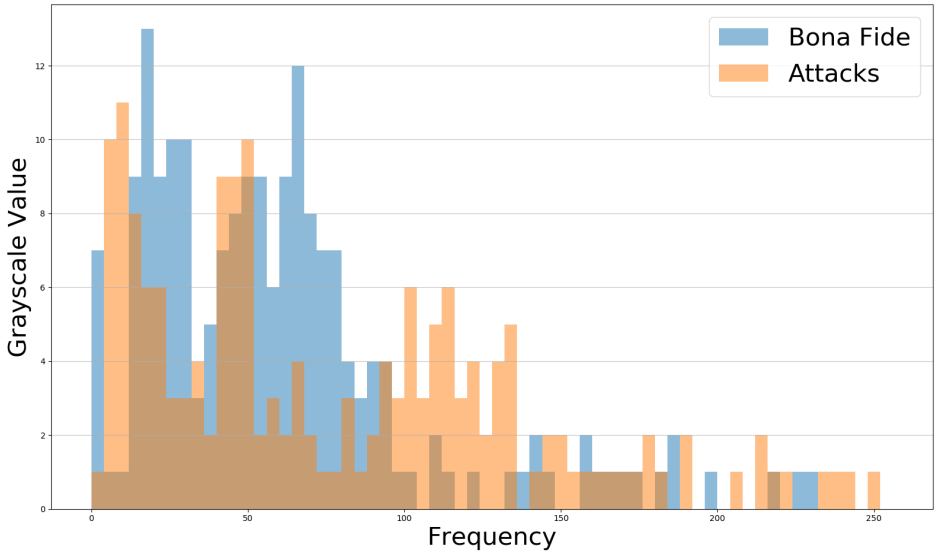


Figure 3.10. – Histogram comparison of Attacks(red) and Bona Fide (blue) in Grayscale.

as bonafide (real) samples (see equation 3.9),

$$BPCER = 1 - \frac{\sum_{i=0}^{B} \zeta}{|B|} \tag{3.8}$$

$$APCER = 1 - \frac{\sum_{i=0}^{A} \xi}{|A|} \tag{3.9}$$

where, |A| and |B| is the total number of presentation attack and real images, respectively. ξ and ζ is 0 if an attack image is classified as real (bonafide), else 1 [ASVN22]. The feature vectors of the histogram approach are created by either concatenating, adding or subtracting the histograms of ultraviolet and visual photo for a total of 9 experiment setups. The different setups are then evaluated based on APCER[1] and BPCER[2]. While the SVM and AdaBoost approaches both yielded usable results (with the AdaBoost approach performing the best). The logistic regression approach was not able to capture the difference of legitimate faces and attacks to an acceptable degree. This is likely due to the high amount of data required to train neural networks in comparison to SVM and AdaBoost. Among the different vector combination approaches, adding and concatenating performed comparatively (adding performing slightly better), while subtracting did not perform as well, likely due to a loss of information when brightness value resulted in zero.

Due to the small amount of data available, the evaluation is performed using a leave-one-out approach. A classifier is trained on using n-1 of the image pairs and then tested on the image pair that was left out. Training/testing is then repeated as many times as image-pairs exist, each time leaving out a different pair to use as the single test case. Since AdaBoost has showed the best results in all scenarios, we only indicate the error rates using that classifier. We were able to achieve a APCER of 0.4% at 1.2% BPCER for the histogram features. Using this feature, we observed false positives (FP) especially in cases using the 2D print attacks. In case of the KP feature

[1]Proportion of attack presentations using the same PAI species incorrectly classified as bona fide presentations in a specific scenario
[2]proportion of bona fide presentations incorrectly classified as presentation attacks in a specific scenario

Table 3.4. – Our Results on the presented Dataset.

Scenario	Histogram		Keypoints		Fused	
	APCER	BPCER	APCER	BPCER	APCER	BPCER
Only Skintype 1-2	0%	0.4%	2.2%	2.45%	0%	0%
Only Skintype 3-4	0%	0.4%	3.3%	3.0%	0%	0%
Only Skintype 5-6	0%	0.4%	3.9%	6.9%	0%	0.2%
All	0.4%	1.2%	4.2%	7.2%	0%	0.2%

vector we achieved 4.2% APCER at 7.2% BPCER while having FP mostly at the attacks using the silicone 3D print and the painted latex masks. By combining both feature vectors into a common one and training them with AdaBoost we reduced the APCER to 0% at 0.2%. This is consistent with our assumption that both properties contain complementary information that together allow a meaningful distinction of the classes.

3.3. Handwriting Identification

When using anonymous offline questionnaires for reviewing services or products it is often not guaranteed that a reviewer does this only once as intended. In this paper an applied combination of different features of handwritten characteristics and its fusion is presented to expose such manipulations. The presented approach covers the aspects of alignment normalization, segmentation, feature extraction, classification and fusion. Nine features from handwritten text, numbers and checkboxes are extracted and used to recognize hand-writer duplicates. The proposed method has been tested on a novel database containing pages of handwritten text produced by 1,734 writers. Furthermore we show that the unified biometric decision using a weighted sum combination rule can significantly improve writer identification performance even on low level features.

In this paper an applied combination of different features of handwritten characteristics and its fusion is presented to expose such manipulations. The presented approach covers the aspects of alignment normalization, segmentation, feature extraction, classification and fusion. Nine features from handwritten text, numbers and checkboxes are extracted and used to recognize hand-writer duplicates. The proposed method has been tested on a novel database containing pages of handwritten text produced by 1,734 writers. Furthermore we show that the unified biometric decision using a weighted sum combination rule can significantly improve writer identification performance even on low level features.

3.3.1. System Overview

The system presented in this paper consists of five parts: alignment normalization, segmentation, feature extraction, classification and fusion. Nine features are extracted from the handwriting elements: 'free text', 'date' and 'checkbox'. The features are mainly based on visible characteristics of the writing, such as: color, slant, word proportions and crosses. Furthermore local interest points in form of SURF Features getting used. The individual steps of the process are shown as program flow chart in Figure 3.11.

(1) Alignment Normalization: The digitized questionnaires show an inconsistent alignment, due to the scanning process or mistake in the printing of the paper questionnaires. In order to correct the handwritten elements with respect to translation, rotation, and uniform scaling, an affine transformation matrix was calculated from the filled in questionnaire and the exemplar questionnaire. The approximated matrix of the affine transformation is

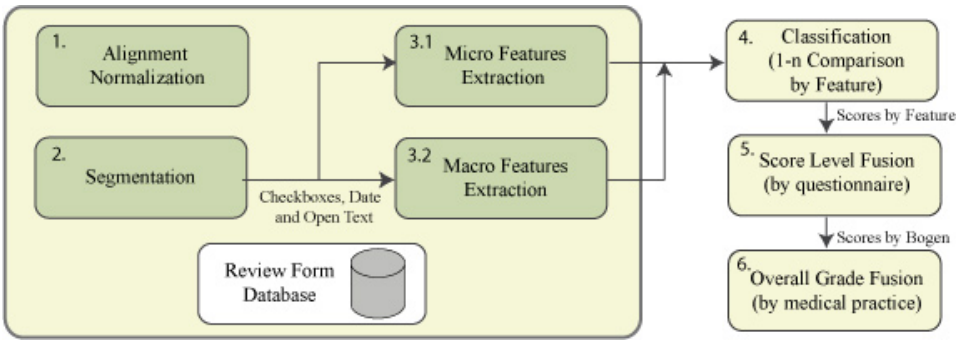


Figure 3.11. – Program flow chart.

used to correct the transformation with up to 5°of freedom in the target image.

(2) Segmentation: Artifacts that may have occurred due to the scanning process get removed using morphological closing. Background colors of the forms get removed using 'color thresholding'. Since areas for handwritten content are highlighted in a rectangular box, through which the contours of the binarized image can be recognized and extracted based on their geometric shape. The 'free text' field is not segmented into text lines or individual words, instead the whole text area is used on the following steps. Where there are checkboxes, the detected rectangles get increased by a fixed pixel value. The surrounding rectangle (resp. underlying line) of the 'free text' and 'date' field get removed. To do this without cutting handwritten text that is crossing those lines, a horizontal and vertical color histogram from a image section of 1*5 pixel is used. When following the lines pixel by pixel the histogram differs where it comes through that a line gets intersected by handwritten text.

(3.1-3.2) Feature Extraction: Different features are extracted from the segmented and classified elements. A more detailed description of each feature follows in chapter 4.

(4) Classification: For each feature type a questionnaire is compared with all other questionnaires (1-n comparison). The calculated similarity scores provide information about the similarity of handwriting in relation to each feature.

(5) Score Level Fusion: For each questionnaire the Similarity Scores are evaluated and merged as a result of the classification of each feature. The result gives information about which handwriting's in the questionnaires are identical.

(6) Overall Grade Fusion: Detected, same handwriting's (step 4) get filtered, based on their frequency of appearance.

3.3.2. Database

The here presented and used database contains 1,734 questionnaires with handwritings from 1,694 different writers. Five subjects filled out multiple (ten) questionnaires that can be used for evaluation of genuine comparisons. The various multi-page questionnaires were digitized with a scanner in high quality and collected as images in

JPEG format. Arrangement and quantity of the handwritten elements depend on the type of questionnaires used. Altogether, there are 34 different types with different arrangements of free text fields, date fields and rating parts. Table 3.5 gives an overview of the used fields and the average amount of analyzed data per questionnaire. In total, the free text fields contain 36,813 words and 213,822 characters.

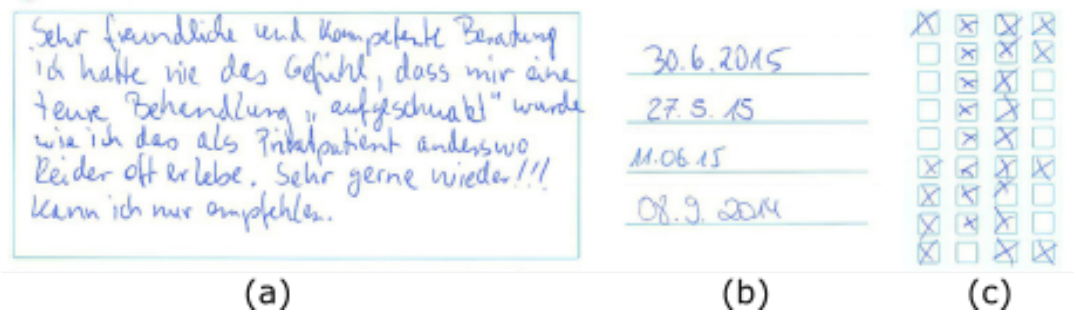


Figure 3.12. – Examples of the analyzed field types: (a) free text field, (b) date field and (c) checkboxes.

Field	Occurrences	Average amount of data
Free text	1,734	24.14 words
Date	1,734	6.24 digits
Checkboxes	12,521	7.23

Table 3.5. – Total number of occurrences and average amount of data.

The language used in the review forms is German. The database contains handwritings of people of different ages and sex. Since the recording process was unconstrained, the handwritings were collected using different pens and blotting pads.

3.3.3. Feature Extraction

3.3.3.1. Writing zones

In this work the 'free text' field was used to compute the height of the ascender and descender parts of single words which are determined by the upper and lower baseline (see image 3.13c). These words get extracted using connencted components, but since not all connected components form complete words, Hough Transformation was employed to connect single letters to words [LSHF95]. The separated words of each free text field are then examined and weighted according to their usability (long words are preferred). For the determination of the two baselines a procedure described by Martıet al [MMB01b] was employed. First, a vertical projection p of the text line image is computed. Then, an ideal histogram $h_i(ub,lb)$ with variable position of the upper baseline ub and lower baseline lb is matched against the projection p by minimizing the square error

$$E(ub,lb) = \sum_i [h_i(ub,lb) - p_i]^2. \tag{3.10}$$

By analyzing the position of the two baselines only words consisting of the upper two zones are selected and their vertical projections are extracted as features.

3.3.3.2. Color Histogram

The calculation of a HSV color space histogram is used for comparing the color of the writing. The width of the histogram represents the distribution of colors while the height shows the frequency density.

3.3.3.3. Line width

The line width is calculated by using the morphological operation 'erosion'. Different structural elements (rectangle, circle, triangle) were iteratively employed to erode the text by a factor of 1. After every single employment the ratio of remaining text elements gets calculated. The rate is summarized by a single value which represents the line thickness.

3.3.3.4. Checkboxes

To our knowledge there has been no research done yet about marks in checkboxes as a feature for handwriting identification. There are three properties that can be analyzed. First, the type of the mark can differ (two separate strokes or one connected line). Second, the length and position of the lines and their relation to one another is specific for an individual writer and third, the order of the strokes differs from writer to writer. To find the nonempty checkboxes all checkboxes were compared to an empty template using a correlation matrix. The box gets eliminated by employing a sliding window so that the crossmark only is left over. Using the Euclidean distance the outer corners of the crossmark are computed and the middle is found by another sliding window. These five keypoints (upper and lower left and right corners and the middle) are stored as a feature vector.

3.3.3.5. Digit Height

The height of the writing gets examined using the date field and here their digits in particular. For cropping the date from the background the morphological operator dilation is used to connect single elements. Then the contours of the text are getting calculated and written on a mask using topological structural analysis. The contours with an higher bounding rectangle than a certain threshold get used as mask on the original image of the date. The slope (see Figure 3.13a) of the date gets removed calculating the minimum-area bounding rectangle of the text within the binarized image. The angle of the rotated bounding box is used to correct rotation of the cropped date. Single digits of the date get extracted using the number of white pixels within a bounding box of connected components in proportion to the width of a line (see section 3.3.3.3). Height and aspect ratio of the digits get calculated using the enclosing bounding box.

3.3.3.6. Slant

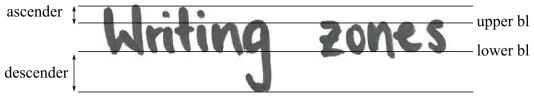
Slant describes the obliqueness of the text or of the individual letters compared to the vertical axis (see Figure 3.13b). The feature can vary both between handwriting's of different people as well as within one word written by one person. For an automated handwriting comparison the calculation of the deviations have been implemented in two different ways. First the approach Bozinovic et. al. [BS89] is used which is based on the calculation of an average angle of almost vertical elements on all text within the free text field. The second implementation [dZ06] was carried out to eliminate identified weaknesses of individuals with untidy handwriting. It calculates means of horizontal and vertical projection histograms of the four 'most usable' classified single words of the 'free text' field (see section 4.1). The individual words are then distorted by a likely minimum to a likely maximum angle. The angle with the highest deflection of the histogram is assumed to be at the correct angle.

3.3.3.7. SURF

The well known SURF features [BTVG06] have shown in many recent papers [WTB14], [JD14] the effectiveness of interest point based methods for writer identification. For each 'free text' field a set of interest points is extracted using the fast hessian detector. The kind of extracted feature points is specified using a library of 120 images. These feature were further processed within a bag of words approach.

(a) Example showing slope.

(b) Example showing slant.



(c) Writing zones.

(d) SURF - before and after applying the algorithm.

Figure 3.13. – Examples of the different extracted features.

3.3.4. Fusion

In this work, the weighted sum combination rule is used to produce a fused unified biometric decision based on the provided decision scores from different features. As a first step for the fusion process, the scores from different biometric sources have to be normalized to a comparable range. Here, the min-max normalization is used and it can be formulated as,

$$S' = \frac{S - min\{S_k\}}{max\{S_k\} - min\{S_k\}} \tag{3.11}$$

Where $min\{S_k\}$ and $max\{S_k\}$ are the minimum and maximum value of scores existing in the training data of the corresponding biometric source and S' is the normalized score. Within the weighted sum score fusion, each biometric source must be given a weight that indicates its relevant effect on the fused decision such that more accurate sources will have larger effect. In this work, the Overlap Deviation Weighting (OLDW) proposed by Damer et. al. [DON14a] is used to control the effect of each biometric source in the final decision. Given the imposter scores S_k^I, the genuine scores S_k^G, the equal error rate EER and the score threshold at the equal error operating point T, the OLDW can be given as:

$$OLD_k = \sigma(\{S_k^I \mid S \geq T\} \cup \{S_k^G \mid S < T\}) \times EER \tag{3.12}$$

$$w_k = \frac{\frac{1}{OLD_k}}{\sum_{k=1}^{N} \frac{1}{OLD_k}} \tag{3.13}$$

Given the calculated weights, the fused score by the weighted sum rule F for N score sources is given as,

$$F = \sum_{k=1}^{N} w_k S_k, k = \{1, \ldots, N\} \tag{3.14}$$

3.3.5. Results

Feature	Field	Classification Metric	EER
Fusion			4.3%
Color Histogram	Open Text	Histogram Correlation	8.2%
SURF	Open Text	Histogram Correlation	10.7%
Slant Hist.	Words	Distance btw. single values	24.7%
Writing Zones	Words	Histogram Correlation	30.1%
Line Width	Open Text	Distance btw. single values	30.2%
Cross Key Points	Checkbox	Distance btw. single values	34.2%
Digit Height	Date Digits	Distance btw. single values	36.1%
Slant	Open Text	Distance btw. single values	37.0%
Digit Aspect Ratio	Date Digit	Distance btw. single values	40.0%

Table 3.6. – Achieved scores of single and fused features.

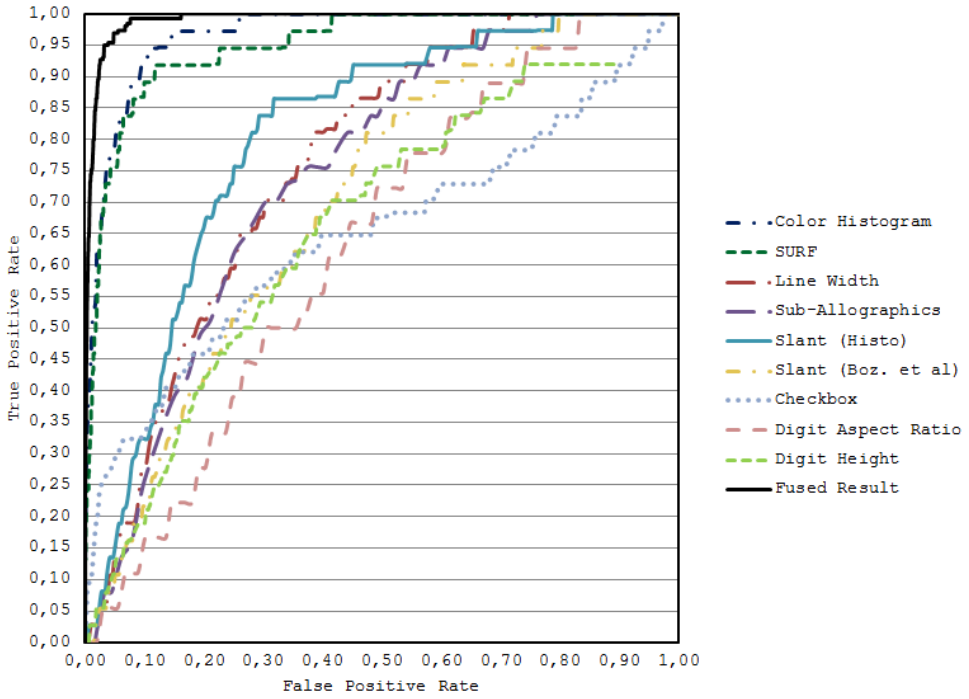


Figure 3.14. – ROC curve showing results by feature type.

In the framework of this development a total of 1,734 questionnaires, based on 34 different questionnaire types were used. In order to evaluate the achieved results four test subjects had each filled in 10 questionnaires (4 times 10 Genuine, 1,694 Imposter). False Rejection Rate (FRR) and False Acceptance Rate (FAR) are calculated for different threshold values (see equations 3.15 and 3.16) where $|A|$ is the total number of real questionnaire and $|B|$ the total number of attacks. The threshold value is the threshold that determines at which similarity value a

detected feature is accepted or rejected. The EER (equal-error-rate) was calculated for each feature separately in order to show their significance (see Table 3.6). HTER is an average of False Rejection Rate (FRR) and False Acceptance Rate (FAR) (see equation 3.17. The EER is a specific value of HTER at which FAR is equal to FRR. This means that at this point the error probability for FAR and FRR is equal.

$$FAR = \frac{\text{False Positives}}{|A|} \tag{3.15}$$

$$FRR = \frac{\text{False Negatives}}{|B|} \tag{3.16}$$

$$HTER = (FAR - FRR)/2 \tag{3.17}$$

In Table 3.6 the results of the different feature methods as well as their classification method are shown. The ROC (Receiver Operating Characteristics) curve in Figure 3.14 shows their characteristic performance under different prioritization of TPR (True-Positive-Rate) and FPR (False-Positive-Rate). With an EER of 8.2% and 10.7% color histogram and SURF features performed best compared to all other features types (see Table 3.6). Local writing characteristics like slant or writing zones are not as distinctive as color or interest point. Although these properties are weaker than others the characteristics represented by them are important factors. While reading the slant of writing from the whole 'open text' field (using the approach by Bozinovic et. al. [BS89]) resulted in a high error rate of 36.97% while vertical histograms of single words showed only 24.7% of EER. A reason for that could be that the slope of writing lines in the 'open text' field affects the distinctness negatively. The analysis of digits and crosses are not based on letters and thus complement the results to multi-modal characteristics. By fusing all scores of the extracted features an EER of 4.33% have been achieved.

3.4. Summary

In this chapter, three methods that identify and improve specific performance parameters of biometric systems were shown. With melanin face pigmentation (MFP), a new modality was presented that has not yet been examined in the literature in the context of biometrics. It was found that the evaluation of this modality is definitely worthwhile (albeit small) and improvements in performance could be observed. This could be shown by the fact that the performance has improved with the addition of the features. This is particularly important given the fact that not all people have MFPs that can be evaluated, and that new MFPs can also appear in the course of aging. The method presented in Section 3.2 for detecting presentation attacks (PAD) in face recognition was able to confirm the importance of this performance parameter. It has been found that when a face image and an image showing MFP in the face are evaluated simultaneously, attacks are detected with a high probability. A novel multispectral face image database comprised of 91 subjects and several face presentation attacks was presented. Various attempts at attack were empirically simulated in experiments and the information content of UV images for the detection of MFP features was confirmed.

In Section 3.3, the problem of manipulating biometric systems by presentation attacks was considered from the point of view of double enrollment in writer recognition. When using offline questionnaires for reviewing services or products, it is often not guaranteed that a reviewer does this only once as intended. A double enrollment check needs to be performed. A combination of different features of handwritten characteristics and its fusion were presented to expose such manipulations. The proposed method has been tested on a novel database containing pages of handwritten text produced by 1,734 writers. Furthermore, it was shown that the unified biometric decision using a weighted sum combination rule can significantly improve writer identification performance even on low level features leading to a duplicate identification rate of 95.67%. In this chapter, the

three important performance parameters: overall accuracy, presentation attack detection, and double enrollment of biometric systems were selected and examined in more detail. Novel solutions to these challenges that can increase the applicability of face and handwriting biometrics in practice were presented.

4. Methods for Enabling Autonomous Entrance Control.

The protection of critical infrastructure is a very current topic and includes both its digital protection through cybersecurity and its physical protection, e.g., through access systems. The separation of individuals when entering buildings is well known, particularly at airports and train stations. Separation systems are used to ensure that only one individual can pass through a designated transit area. Depending on the location, different physical appliances such as turnstiles or airlocks are used. A special industrial application is so-called mantrap portals, which consist of a room with two doors. A person enters through the first door and the second door gives the person access to the protected area. The special thing about these portals is that they only release the second door when various requirements are met. These are usually:

1. The person belongs to the group with access authorization
2. The person is alone in the portal

There are numerous ways to ensure that the first condition is met, from items that indicate possession such as magnetic cards and RFID chips to knowledge-based authentication using PIN or passwords to biometric systems. The second condition is much more difficult to meet with an autonomous IT system. The reasons for this are that the system has to detect that there is more than one person at a designated area, even if the person is carrying items with them (such as vacuum cleaners and cash boxes) or changes their appearance in any way. Already registered people could be wearing glasses, caps/hats, backpacks, or jackets/coats. It also needs to be taken into account that people deliberately try to trick the system by using certain devices, or people perform tailgating and piggybacking in order to spoof the system. For this reason, the goal of this chapter is to answer:

Research Question 3: Which technologies are suitable for autonomously detecting piggybacking and tailgating when accessing restricted areas?

Solutions currently used in industry are not tamper-proof and consist of simple sensors that can be overcome with little prior knowledge. Unfortunately, there are few published studies in this area, so the actual security of such systems can hardly be evaluated and compared with others. It is also not clear which sensor technology contains the most useful information. The challenge is therefore not only to carry out a screening of suitable technologies, but also to develop a test procedure and database for how these technologies can be independently checked for their security against being defeated. In Section 4.1, the scenario is explained in more detail and the technical properties of the mantrap portal are discussed. The circumstances and prerequisites with which the data was collected in order to test the technologies are embedded into the following Sections 4.2-4.6. The first Section will examine thermal images to detect humans based on their body heat. It is based on the analysis of false-color images of an infra-red thermal camera mounted on the ceiling and presents a novel method based on the image variance. The second analyzed approach (see Section 4.3 evaluates images with additional depth information. So-called RGB-D images are false-color images that use a color to indicate the distance between a pixel and the sensor. The third approach is based on a series of 21 normal images without additional information captured over a period of 3-4 seconds. The method presented uses optical flow, which makes movements in the image evaluable.

Instead of visually estimating the position of humans, the fourth method (see Section 4.5measured the pervasive capacitance of feet in the transit space to detect tailgating attacks. Suitable sensing techniques and sensor-grid layout are to be used for that application as well as machine learning techniques for classification of the sensor's feature vector. In Section 4.6, sensors installed in the floor as well as camera shots taken from above are used. A multi-sensor solution for verifying the number of persons that stand within a defined transit area is presented. In the Summary 4.7, all approaches are analyzed to determine whether they reliably recognize piggybacking and tailgating autonomously, and whether the methods examined can therefore be applied industrially.

4.1. Mantrap Portals

The used mantrap portal has a square area of 900x900mm. The height of 2130mm is adapted to the internal installation of the office buildings. The system has two doors attached opposite to each other (see Fig. 4.2) with central frosted glass interior trim (see Figure 4.1 for an exemplary model). The structure is closed from above with an LED light at its ceiling mounted for illumination. The used mantrap portal (see Fig. 4.2) has entrance and exit doors. It is built regarding to the protection classes 'WK3' defined in EN 1627 [EN109] standard for security and burglary resistance and BR 3-S defined in EN 1063 [EN101] for resistance against bullet attack. Subjects enter the mantrap portal, close the door and the system then verifies that only one person is in the portal. A verification weather a person is authorized to pass through this area is usually additionally applied but will not be part of this study. After successful authentication and positive separation the second door gets unlocked and the subject enters the secure area.

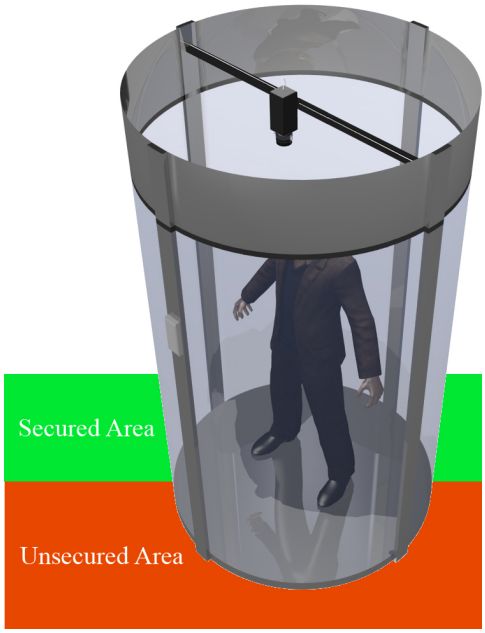


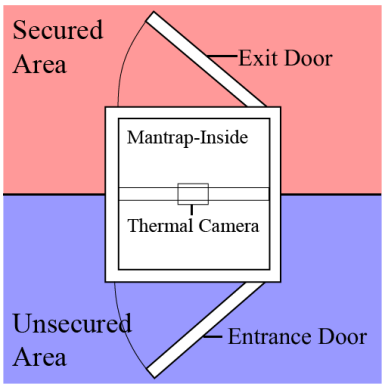
Figure 4.1. – Exemplary model of a Mantrap Portal

4.2. Thermal Imaging

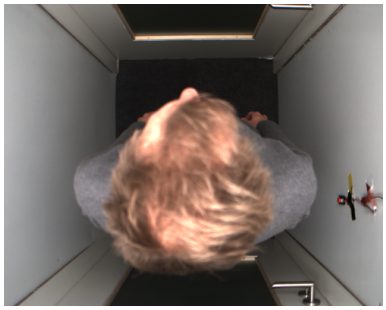
The here presented technical approach uses thermal images to detect humans based on their body heat. It explores the discriminability between human body heat and heat from other non-living objects in thermal images. It is based on the analysis of false-colour images of an infra-red thermal camera mounted on the ceiling. Particular attention was paid to the compromising of body features as defined in VdS3112 [VdS09] set to protect against sabotage. A special focus was on the behavior of the system placed under attack when an intruder tries to overcome the system. The performance was evaluated in empirical testing with a test group, selected according to their physical characteristics. The test scenarios cover changing appearances of individuals and possibly carried objects into the mantrap. The results are presented as Equal Error Rates (EER) and Receiver Operating Characteristics (ROC) curves. Furthermore, impacts of environmental factors such as light and temperature were examined to determine the performance output.

4.2.1. Aspects of the applied Approach

The used images are captured from centered position on the ceiling construction using a thermal imaging camera of the type Flir AX8 [FS16] with a field of view of 48°× 37°with a fixed focus. The focal length of the camera is 24,6mm and the resolution of the IR sensor is 80x60 pixels with a temperature measurement range of -10°C to + 150°C.



(a)



(b)

Figure 4.2. – (a) Mantrap (b) Image from Top View

The camera offers an adaptive adjustment of the analysed temperature range, with which changes of temperature in the camera field of view are highlighted. The colour values may vary depending upon the mean temperature in the target area. Areas below the mean temperature are highlighted blue, higher temperatures shown in yellow to white.

4.2.2. Enrolment

Over a period of 3-4 seconds 120 false-colour images of the thermal imager are read and analysed. The classification of a pixel as belonging to a living or non-living object, depends on the colour value of the pixel. To do this, the image is converted into the HSV colour space and allocated to a fixed value range of the H-dimension as a

threshold for "living" objects. For every frame, the number of these pixels is counted and the average calculated. The measurement period also determines the variance which will be added to the pixel average. The calculation of the variance also includes the min-max values of the pixels. Strong movements such as bending or jumping lead to large differences in the thermal image with regards to the sum of the pixels which correspond with the value range of a 'living' object, thus increases the variance and the average value. Additionally a threshold got added that relies on the desired performance of the system (False Positive Rate 'TPR' to True Positive Rate 'TPR'). This reference value was calculated for every subject.

The reference value κ is computed as follows: Equation 4.1 applies fixed-level thresholding to the nth single-channel image $src^n(x,y),\ n = 0,\ldots,120$. The threshold for this operation is ε.

$$dst^n(x,y) = \begin{cases} src^n(x,y) & if\ src^n(x,y) > \varepsilon \\ 0 & otherwise \end{cases} \tag{4.1}$$

Then the mean value of the thresholded images $dst^n(x,y)$ is computed in equation 4.2).

$$\overline{dst}(x,y) = \frac{1}{n}\sum_{i=1}^{n} dst^i(x,y) \tag{4.2}$$

Afterwards the variance is computed as shown in equation 4.3.

$$\mathrm{Var} = \frac{1}{n-1}\sum_{i=1}^{n}(dst^i(x,y) - \overline{dst}(x,y))^2 \tag{4.3}$$

Now the reference value κ is calculated with equation 4.4.

$$\kappa = \mathrm{Var} + \overline{dst} + \alpha \tag{4.4}$$

Hereby α is a threshold that relies on the desired performance of the system (FPR to TPR).

4.2.2.1. Verification

During verification, a value is calculated and compared with the reference value stored during enrollment. The same amount of thermal images are read out over a period of 3-4 seconds. The mean value of the pixel quantity is determined and the variance is added. Verification is successful if the calculated value during verification is lower than the reference value.

4.2.2.2. Test Scenarios

The performance of the system was examined by verifying one person alone and with objects. These objects got selected as they are often carried in security areas (see Table 4.1). In Figure 4.3 a subject with a bucket of warm water is shown as thermal image.

To determine the robustness of the created system against attacks in a real-life scenario, attacking schemes under which two people enter the mantrap were defined. It was assumed that one person had a basic knowledge of the function of the system. In the context of the test, this person represented a "person with authorised access". The second person demonstrated a variety of levels of understanding about the functionality of the system and acted as the "attacker" trying to gain unauthorised access. The seven study scenarios and the used aids are described in Table 4.2 and 4.3. These also include attempted "attacks" with aids as defined in Table 3.

Table 4.1. – Permitted Objects

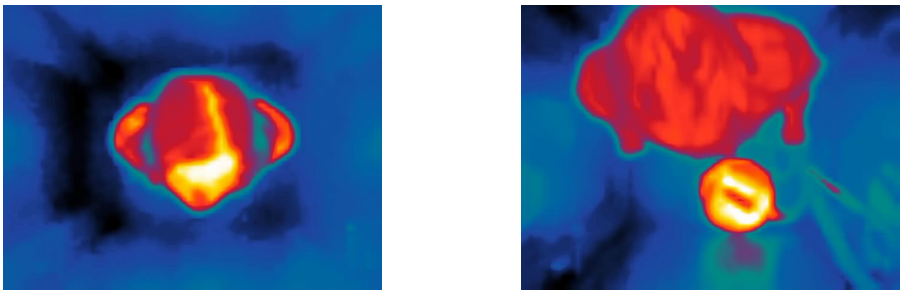
Type	Properties
Plastic bucket	Rectangular form, 10L Capacity, Filled with about 8 litres of water at 50°C
Vacuum cleaner	Brand: Fakir, Model: S20L, Surface temperature about 45°C
Parcel	Material: cardboard, Partially covered with adhesive tape, Dimensions: 98 x 40 x 30 cm (WxDxH)

Table 4.2. – Attack Scenarios

#	Knowledge (attacker)	Aids utilised
1	None	None
2	The "attacker" is aware that a Separation of Individuals is performed however may not be aware which technical Method is applied to do this.	None
3	The "attacker" is aware that a separation of Individuals is performed and is also aware that thermal Imaging is used.	None
4	As Scheme 3	Motorbike helmet
5	As Scheme 3	I-Pad
6	As Scheme 3	Mirror
7	As Scheme 3	Sheet of aluminium

Table 4.3. – Aids utilised by Attacker

Type	Properties
Motorbike helmet	Design: Flip-up helmet, Material: Polycarbonate, Size: XL
I-Pad	Dimensions: 24.28 x 18.97 x 1.34 cm
Sheet of aluminium	Material: Two-layered Polyester Film (PET) according to DIN 13232, Features: Extremely Thin, Highly Reflective Dimensions: 160x210cm, 12μ
Mirror	Finish: Circular, Diameter: 50 cm



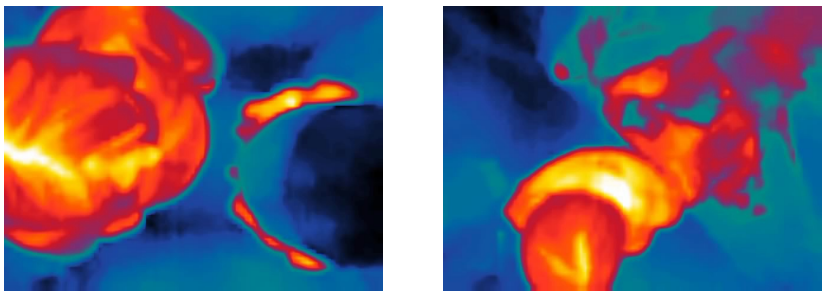
(a) (b)

Figure 4.3. – (a) Thermal Image during Enrolment without an Object (b) Thermal Image during Enrolment with filled Plastic Bucket

4.2.3. Evaluation

As suggested the factors height and stature of a person have an impact on the performance of the system. Subjects have been selected in line with defined physical properties in height and weight. There are two subgroups that have been divided according to BMI (body mass index) in lean and normal. The height of a subject formed three more groups (small, medium, tall) so that there where six different body-types. Because the test subjects may negatively affect the evaluation process and/or the evaluation of the measurement by their behaviour, the test subjects were informed about the procedure and guided by the administrators. The test subjects positioned themselves and the object wherever they chose in the mantrap. After approval by a laboratory employee, the subjects started the separation themselves using a push button mounted in the mantrap. As the test subjects were required to wear different clothes in order to carry out a realistic measurement, the data collection took place on a different day to the verification. During the measurement, test subjects took on different roles depending on the items carried (see Table 1). Reference values were collected from a total of 12 test subjects. The mantrap allowed no distractions in the form of sound or light. The measurements were carried out over a period of three weeks, at different time of a day.

For the data acquisition in the attack schemes, the subjects entered the mantrap in pairs which reflected all possible combinations of physical characteristics. Each person played the two roles of “authorized person" and "attacker” alternately.



(a) (b)

Figure 4.4. – (a) Attack Scenario 4 (b) Attack Scenario 7

4.2.4. Results

4.2.4.1. Attack Scenarios

The first attack scenarios resulted in a very low FPR (False Positive Rate) of 0% in Scenario 1 and 3% in Scenarios 2 and 3 (see Table 4.4). This means that in these scenarios (see Table 4.2), there were none or very few cases of the system being deceived. When the threshold value α was increased by 5%, the FPR rate was at 6%. One possible reason for this behaviour was the items carried in scenarios 4-7, which were deliberately used to manipulate the system. An attacker without an aid has little chance of deceiving the system. Scenarios 4 to 7 (see Table 4.2) with helmet, aluminium sheet, mirror and tablet PC had a considerably more negative effect on the results. Figure 4.2.3 shows examples of scenario 4 and 7 as thermal images. The success rate of these targeted attacks carried out with both knowledge on the part of the attacker and aids was 57%. The main reason for the poor performance of the mirror attack scenario is that it allows all of the attacker's body heat to be removed from the image. If you hold the toy correctly, you can virtually make yourself invisible. Wearing a helmet or holding a tablet in front of you has a similar effect, although the body and shoulders usually cannot be completely covered. The aluminum foil was also able to keep out a lot of heat, but it also reflected heat from the person with access, so the results are slightly better. There was no correlation between the body size or stature of the attacker and their chance of success. The different results of some groups, however, suggest that the difference in posture and position of individuals during the enrolment have an impact. The ROC curve (see Fig. 4.5) shows the archived FPR in relation to the TPR (True Positive Rate) of all verification attempts.

Table 4.4. – False Positive Rates By Scenario

Scenario	TH $\alpha = 0$	TH $\alpha = 5$	TH $\alpha = 10$
1 No previous Knowledge	0%	6%	8%
2 Basic previous Knowledge	3%	6%	11%
3 Adv. previous Knowledge	3%	6%	6%
4 Helmet	14%	22%	36%
5 I-Pad	22%	31%	39%
6 Mirror	39%	47%	56%
7 Sheet of Aluminium	19%	22%	28%

4.2.4.2. Verification of Authorized Subjects

The TPR was calculated from the number of rejected verification attempts of authorized persons and the total number of attempts. Scenarios both with and without objects known to the system were tested. Readings were produced for all 12 subjects involved in the enrolment phase. These scenarios have differing impacts on the TPR. It is thought that the behaviour of different test subjects during the measurement is in part responsible for this. The objects also exert an influence upon the results. Thus, the TPR is lower in the scenario with a parcel than in the scenario with a bucket which has the highest rate.

4.2.4.3. System Performance

The evaluation of the test results (see Fig. 4.5) as well as of the various threshold changes show that the system has an overall EER (Equal Error Rate) of 20.2% when α is 5% of the calculated variance. When looking only at the attack scenarios 1-3, which were carried out without any aids, an EER of only 7.9% have been archived.

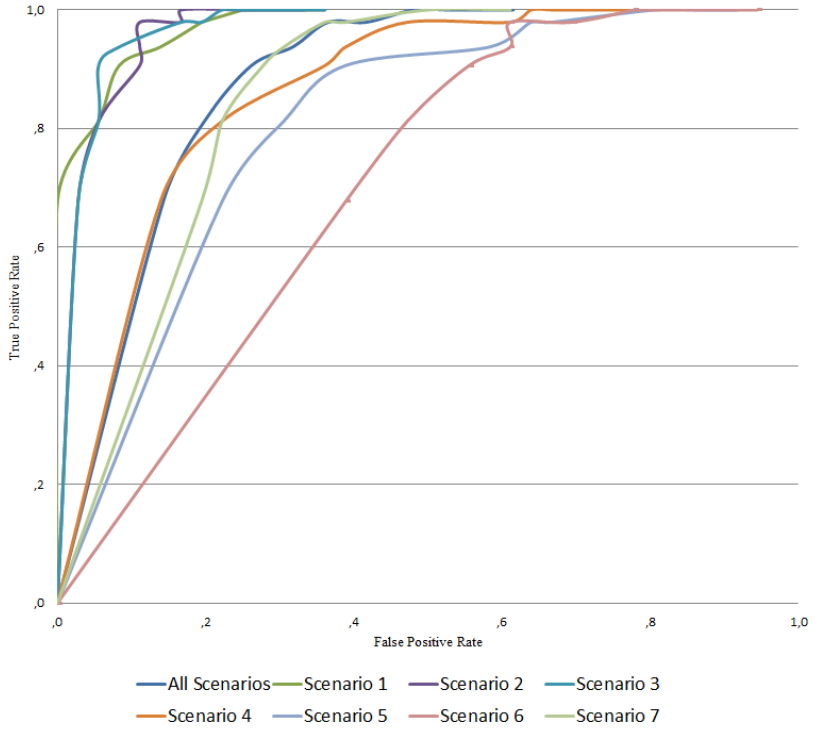


Figure 4.5. – ROC-Curve showing the System Performance for different Attack Scenarios.

Existing methods have not been evaluated under attack and can therefore badly compared with the here archived results. When used in high security access control, the use of a threshold resulting in equally weighted failure rates might not be useful. Instead a lower TPR might be tolerable (what means that a authorized subject might need more attempts) in order to guarantee a lower FPR and therefore more security.

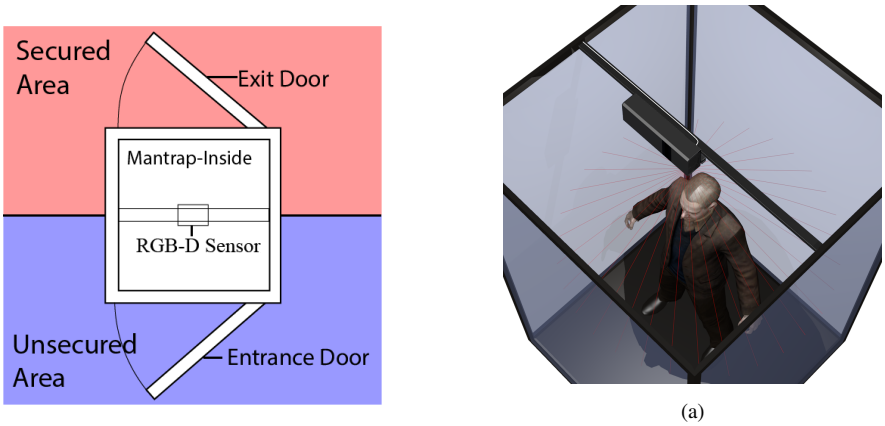
4.3. RGB-D Imaging

Automatic entrance systems are increasingly gaining importance to guarantee security in e.g. critical infrastructure. A pipeline is presented which verifies that only a single, authorized subject can enter a secured area. Verification scenarios are carried out by using a set of RGB-D images. Features, invariant to rotation and pose are used and classified by different metrics to be applied in real-time. The performance was evaluated by using scenarios in which the system was attacked by a second subject. The results show that the presented approach outer-performs competitive methods.

4.3.1. The Evaluation Environment

As a capturing device a RGB-D camera of the type Microsoft Kinect 2.0 [Cor16] was centered mounted on the ceiling (see Figure 4.6). The Kinect 2.0 ToF (time-of-flight) sensor has a field of view of 70.6°x 60°[Sme16] resulting in an average of about 7 x 7 pixels per degree. The used time of flight method measures the time-of-

flight of a light signal between the camera and the subject for each point of the image. The resolution of the resulting depth image is 512 x 424 pixels. Our experiments showed that objects in a distance from 400mm to 8000mm from the sensor can be read.



(a)

Figure 4.6. – (a) Mantrap functionality (b) 3D Model Mantrap Portal

Due to the dimensions of the camera itself and the maximal height of the ceiling was the camera mounted at 2300mm height above the mantrap. The camera generates a point cloud with the geometric parameter x,y and z. Objects closer than 400mm to the sensor result in missing points for that objects. The reliability of the proposed approaches is therefore reduced to subjects not higher than 1.9m.

4.3.1.1. Examined Scenarios

The presented approach is not based on biometrics as the examined modality does not match biometric criteria in terms of uniqueness. Generally biometric systems can be either a verification system or an identification system. The differences between these two types of systems differ in how quickly a system operates and how accurate it is. In this work two different approaches targeting verification with the claim (VIC) of an identity and verification without a claimed identity (VWIC) are presented.
The verification of a subject, that claimed their identity, is based basically a comparison of the captured characteristic (sample) with its reference. In contrast, verification of a subject, without claimed identity, is a comparison between the captured characteristics (sample) with a model of the two classes 'single subject' and 'multiple subjects'. VWIC in this work can be seen as closed set identification as there is only a binary classification between the classes 'one subject' and 'multiple subjects'. The VIC scenario reaches usually lower error rates as a sample of the captured subject is compared only against the references of the claimed identity (1:1 comparison). In our use-cases, the verification sample (containing either one or two subjects or a subject with objects) is compared against a previously enrolled one. In a VWID scenario, a sample is compared against all templates in the database showing a single subject (1:n comparison).
Authorized subjects sometimes need to pass the portal carrying different objects like vacuum cleaners or cash boxes. These objects differ in their appearance and therefore it is not possible to enroll them individually. To be able to test the system performance for those special cases, the performance of the system was examined by including cases of a single subject carrying objects. These objects got selected as they are often carried in security areas (see Table 4.1).

To determine the robustness of the created system against attacks in a real-life scenarios, attacking schemes under which two people enter the mantrap were defined. It was assumed that one subject has a basic knowledge of the function of the system. In the context of the test, this subject represented a "subject with authorized access". The second subject demonstrated a variety of levels of understanding about the functionality of the system and acted as the "attacker" trying to gain unauthorized access. Compared to our previous work in which thermal imaging was used [SHK16] the scenario using a I-Pad was rejected. The six study scenarios and the used aids are described in Table 4.2 and 4.3.

4.3.1.2. Data Acquisition

The participating subjects were selected in line with defined physical properties in height and weight in order to cover a variety of body shapes. There were two subgroups that have been divided according to BMI (body mass index) in lean and normal. The height of a subject formed three more groups (small, medium, tall) so that there where six different body-types. Because the test subjects may negatively affect the evaluation process and/or the evaluation of the measurement by their behavior, the test subjects were informed about the procedure and guided by the laboratory employee. The test subjects were instructed to position themselves in the middle of the mantrap portal and to turn to a specified side of the room. They were furthermore told to focus a marked point on one side of the portal's wall. The recording process was started in agreement with the test subject by an laboratory employee. The data collection (enrollment, genuine and imposter) took place on different days. A total of 12 subjects were enrolled in the data acquisition process. For the genuine data collection each test subject was recorded with and without carrying items (see Table 4.1). In case of the attack scenarios (imposter), the subjects were divided into groups of 6 subjects were each group played either the role of an "authorized person" or "attacker" alternately (see Table 4.2 and Figure 4.8).

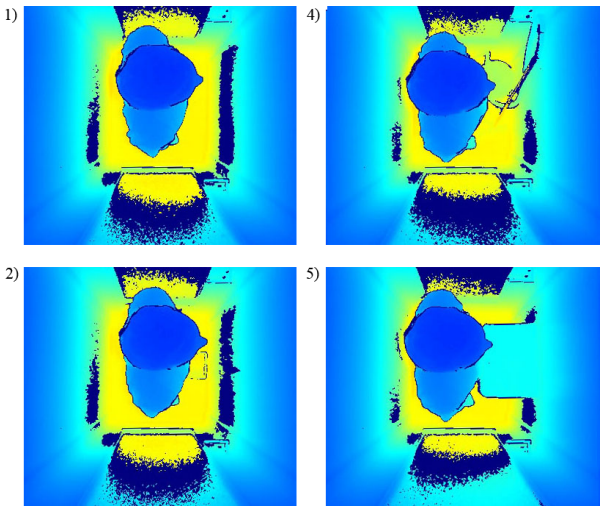


Figure 4.7. – 'Subject A' Genuine Verification Attempts with- and without Objects.

The used mantrap portal allowed no distractions in the form of sound or light. The measurements were carried out over a period of four weeks, at different time of a day. For the data acquisition in the attack schemes, the subjects entered the mantrap in pairs which reflected all possible combinations of physical characteristics. A total of 21 point clouds over a time period of approx. 3-4 seconds were collected in every recording.

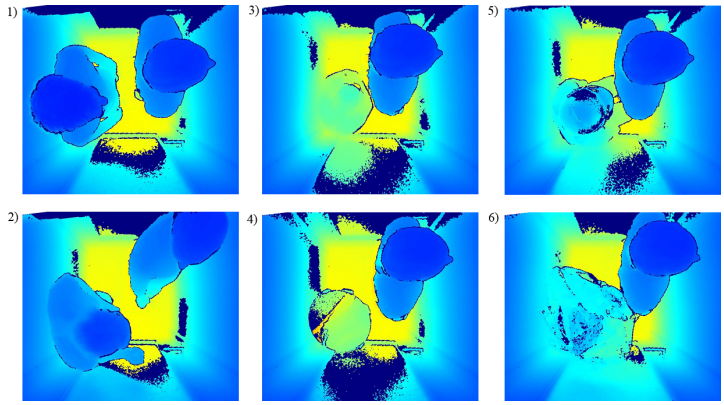
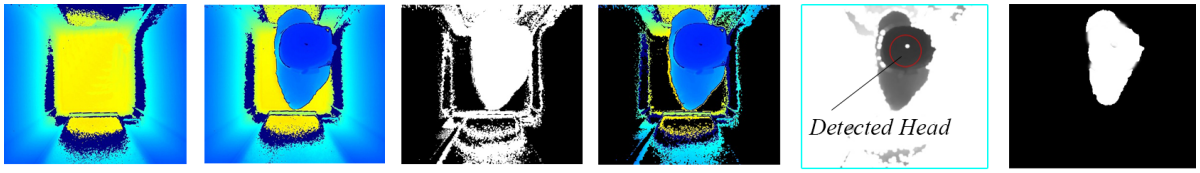


Figure 4.8. – 'Subject A' Imposter Verification Attempts with different Attackers.



Im_b *Background Image* | Im_s *Acquired Sample* | Im_c *Foreground Mask* | Im_d *Background Subtr.* | Im_f *Head and Shoulder Detection* | Im_g *Body Contour Mask*

Figure 4.9. – Images showing Pre-processing and generation of Shape- and Background Model.

The point clouds from the Kinect RGB-D camera were converted into 3-channel 2D matrices by applying a color-map to the z value of the point cloud. For pixel with no corresponding point in the cloud the color-value 0 was applied. For further processing the images were converted into the HSV (Hue-Saturation-Value) color space.

There were also background images of the empty portal acquired (see image $\{Im_b\}$ in Figure **??**) that showed some differences compared to the images during acquisition (see image B in Figure 5.11). Especially in corners and at the doors of the portal, noise was given (dark blue areas indicates 'no depth value'). This could be reasoned by reflections from the walls, the glass components in the door areas or the accuracy of the time of flight sensor. Comparing a depth value of a specific point between the background image and the image when a subject entered the room, the value could either: a) be the same b) change slightly or c) disappear. As humans have different shapes and can never stand completely still, these noise was one of the challenges to deal with.

4.3.2. The Proposed Methods

In order to verify a image sample, following pipeline shown in Figure 4.10 is presented which is based on the idea of using a subject shape model and a background model for comparison.

As at least one subject is present in the image sample, the subject shape model can be used for comparison with the previously enrolled reference model. The background model is needed to ensure that the remaining part of the portal is the same as during enrollment. Following steps describe how the head area was detected. The used parameter were calculated by testing iteratively with the acquired images and shown in Table 4.5.

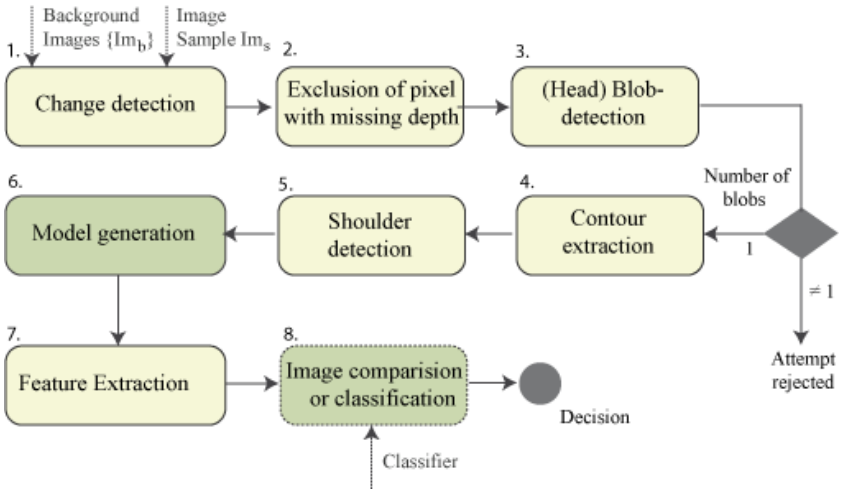


Figure 4.10. – Used Pipeline.

4.3.2.1. Foreground selection

The selection of foreground can be performed on the base of appearance or depth. Appearance based methods are affected by shadows, whereas depth methods are imprecise because of foreground fattening [SSZ01]. An adaptive Gaussian mixture-based background/foreground segmentation method [Ziv04] was used to exclude the parts from the sample image that have the same texture as the background image. It allows dynamic background elements to be modeled through color intensities at individual pixel locations using a mixture of Gaussian probability density functions. That method has the advantage that it is more robust against focus problems with shadows and multi-modal background regions than non-adaptive methods using just a single reference image. Nevertheless, the background images used for training the model, were acquired in a separate recording of the empty portal and then applied on the algorithm for all samples individually. Parts of the samples which belong to the portal and have not been changed during the recording were therefore extensively excluded.

4.3.2.2. Missing depth exclusion

As discussed in the data acquisition chapter of this work, some random noise was given in all images due to the scenery. For this reason the foreground mask still (see Figure 4.9 Im_c) considers parts of the background to be foreground. The pixel of an image where no information was given either on the sample Im_s or the background images Im_b were therefore excluded by applying pixel wise the color value 0 to them. The results of this pre-processing step are shown in Figure 4.9 Im_f

4.3.2.3. Blob detection

The goal in the blob detection step is to find the circle like head area. A head Blob is considered to be a group of connected pixels in the image that form a circular like shape. The blob detection has follows steps:

1. First, the image is converted into gray scale. The sample is then converted into various binary images applying lower and upper color thresholds. The distance between neighboring thresholds defines the amount of binary images.

2. Structural component analysis is used create connected components from the grouped white pixel of the binary images. Filtering of the detected blobs detects only circular 'head alike' blobs of a certain size. Following formula is used to filter blobs for their min- and max circularity:

$$\frac{4\pi \cdot area}{perimeter \cdot perimeter} \tag{4.5}$$

3. The centers of the grouped blobs is computed and blobs located closer than a minimal distance are merged.
4. The center points and radius of the merged blobs are computed and used in further segmentation of the shoulder area (see Figure 4.9 Im_f for an illustration of the detected head-blob).

Table 4.5. – Parameters for (Head) Blob Detection.

Name of Parameter	Value
Minimal color threshold	50 (color in gray scale)
Maximal color threshold	130 (color in gray scale)
Minimal areas of blob	2300 pixel
Maximal areas of blob	15000 pixel
Minimal circularity	0.1
Minimal convexity	0.87

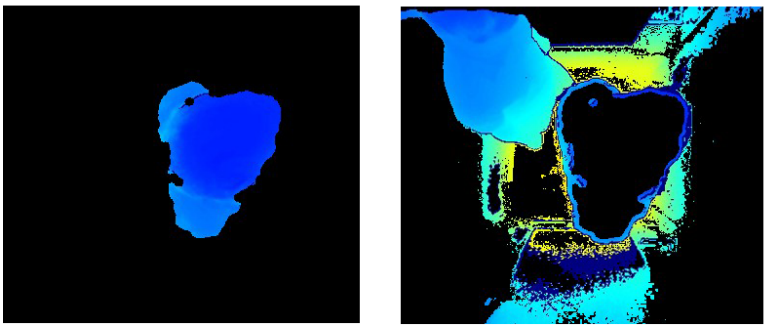
If the number of detected blobs is unlike 'one', the verification attempt is counted as rejected. Rejected samples with a single subject on the image were counted as False-Negative (FN) and rejected samples with more than one subject as True-Negative (TN).

4.3.2.4. Contour extraction

In the contour extraction step, the head area is selected and its average color is calculated. For selecting the head area, the given radius and the center-point were used. The average color values for this area were used to calculate a range for the shoulder area.

4.3.2.5. Shoulder detection

The range for detecting the head and shoulder area was defined using a upper and lower color-offset. The lower boundary was calculated using the head average color, minus a constant value. The upper boundary was calculated by adding a constant value to the head average height. The constants where defined by manually testing with the collected data. Morphological closing was applied using a ellipse element of size 5x5 pixel to eliminate noise in the image. The image was then binarized and structural component analysis applied in order to receive connected components. The connected components were approximated using a distance between the original curve and its approximation. The Douglas-Peucker algorithm was used for the approximation, so that elements with a maximum distance of 3 pixel were closed. For each component a circle of the minimal enclosing area was created. The distance between the center-point of the enclosing circle center and the head center was calculated using the euclidean distance. The distance and the size of the components were used to filter the components.



Im_{sm} *Shape Model* Im_{bm} *Background Model*

Figure 4.11. – Shape and background models for a genuine attempt.

4.3.2.6. Shape and background model generation

The area of the components calculated in the head and shoulder detection steps were used to calculate a binary mask (see Figure 4.9 Im_g). The mask was used to extract a color body shape from the pre-processed image Im_d. The remaining parts of image Im_d were used as background mask (see Figures 4.11 Im_sm and Im_bm).

4.3.2.7. Feature extraction

A color histogram, using the HSV color model was extracted from the background model using 180 bins. The value of each bin represents the appearance of the color in the image. These kind of feature is used because it is rotation invariant and takes the body height and size of the shape into account.

4.3.2.8. Image comparison or classification

In the verification with identity claim scenario, the feature vectors were compared directly with the enrolled template of the same subject. The comparison metric correlation performed best, compared to other metrics as they are Chi-Square, Intersection and Bhattacharyya distance. A verification attempt is successful if the correlation score between the sample feature vector and the enrolled template is below a certain threshold (see chapter results). In the verification without identity claim scenario a classification of the feature vector is performed using the machine learning technique SVM. In this scenario a classifier was trained using a number of images of both cases (also called 'classes') showing one or two subjects. Training images from a selection of 'genuine verification attempts' (see Figure 4.7) and imposter verification attempts (see Figure 4.8) and 4.9 were used.

4.3.3. Results

An Evaluation of the blob detection was performed in order to define the accuracy of this pre-processing step. On all images with one head, a detection accuracy of 98.4% was achieved.

In the case of verification with identity claim, the first two attack scenarios without additional aids were recognized completely (see 4.6). While the third attack scenario, where the attacker had advanced knowledge

about the system resulted in an EER of 12%. In many images of that attack scenario, the attacker was lying flat on the ground or hided himself below the access allowed person (see Figure 4.8 image 3 for an example). In that cases, the tolerance of the depth sensor was too high to differ between ground and attacker or access allowed subject and attacker. In some cases the attacker could also put himself in position where its body and the noise of RGB-D image were overlapping. The attack scenarios with aids resulted in less than 5% EER. This is reasoned because the depth images that contain additional objects besides the attacker body, make a bigger differences in the depth when comparing with the reference image or model. As a result a attacker cannot deceive the system by using these additional aids on the contrary of the method using thermal imaging.

In the scenarios where no identity was claimed, the error rates in general increased. The main reason for that was the big variation of body shapes within our test-group. Besides that, the classifier lost some of its accuracy because of the noise given in the images. The scenario in which a mirror was used to spoof the system resulted in lower error rates, which shows that the RGB-D is not that vulnerable to those attacks such as thermal imaging.

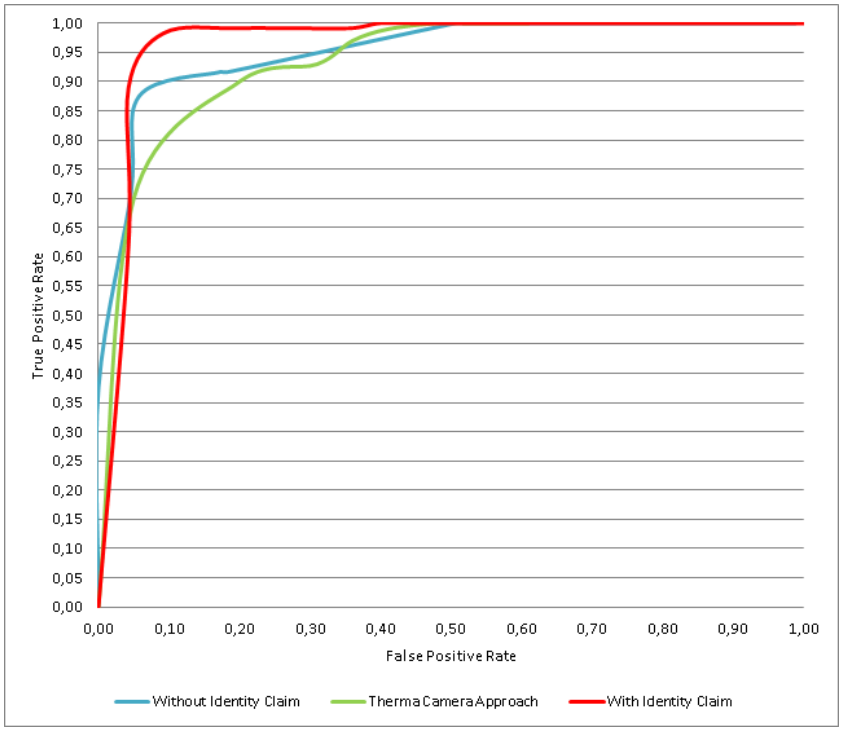


Figure 4.12. – Averaged Results for Verification with/without Claim of an Identity against the Baseline.

The size of carried objects have an impact on the system, at is was observed that the scenario when subjects where carrying a parcel archived the lowed accuracy in the genuine comparisons. There was no correlation between the body size or stature of the attacker and their chance of success. The ROC curve (see Fig. 4.12) shows the archived FPR (False Positive Rate) in relation to the TPR (True Positive Rate) of all verification attepts with and without identity claim, for all different scenarios. The TPR was calculated from the number of accepted or rejected attempts of authorized subjects and the total number of attempts. The tests resulted in an overall EER (Equal Error Rate) of 5% for the verification with identity claim scenario and 11% for the verification without identity claim scenario (see Figure 4.16). When looking only at the attack scenarios 1-3, which were carried out without any aids, resulted in higher EER's of 4% resp. 15% in the scenario without identity claim. In contrast,

the scenarios 4-6 in which additional aids were used achieved equal error rates by an average of 2% resp. 9%. Compared with the previously examined approach using thermal imaging [SHK16] achieves the here presented solution lower error rates in most of the attack scenarios. When used operationally in physical access control, the use of a threshold resulting in equally weighted error rates might be not recommended. Instead a lower TPR might be tolerable, which would result in a less convenient system but would guarantee a lower FPR and therefore higher security.

Table 4.6. – Equal Error Rates By Scenario

Scenario	Verification w. Identity Claim	Verification w/o Identity Claim	Thermal
All	5%	11%	22%
1 No Knowledge	0%	11%	6%
2 Basic Knowledge	0%	12%	6%
3 Adv. Knowledge	12%	22%	6%
4 Helmet	4%	7%	26%
5 Mirror	0%	7%	56%
6 Sheet of Aluminium	1%	12%	28%

4.4. Optical Flow

Unstaffed access control portals are increasingly used in high security areas. Existing systems require expensive hardware, or are sensible to changing environmental conditions. We present a single camera system for a mantrap that verifies that only one individual is in the designated transit area. Our novel approach combines optical flow and machine learning classification. A database was created that consists of images of attempted attacks and regular verification. The results show that our approach provides competitive results and outperforms detection rates in several attack scenarios.

4.4.1. The Evaluation Environment

In the center of the ceiling, a color camera manufactured by Point Grey with CMOS sensor was installed. The resolution of the acquired images was 1280x1024 pixels at 8 bit RGB.

Due to the dimensions of the camera itself and the maximum height of the ceiling, the camera was mounted at 2300 mm.

4.4.1.1. Examined Scenarios

We evaluated our approach using biometric evaluation metrics despite it not being based on biometrics. This is due to the human¬-centered binary classification approach. Biometric systems are split into verification and identification. In this work we present a verification use-case in which the identity of the subjects is not claimed.

In this work we can see the verification case as closed set, because it only differentiates between 'one subject' and 'more subjects'. In our use-cases, the verification sample (containing either one or two subjects or a subject with objects) was classified using a machine-learning technique. A sample was compared to a model that was trained using all entries of the database showing attacks (class 1) or single subjects. Authorized subjects need to

pass through the mantrap portal carrying different objects. These include vacuum cleaners and cash boxes. As there are a lot of variations of these objects, they cannot be enrolled individually. To test the system performance in these cases, the performance was tested by verifying one subject alone and with objects.

We selected these objects as they are often carried in security areas (see Table: 4.1).

The robustness of the created system against attacks was calculated by defining different attack schemes. The schemes involved two people who try to pass through the mantrap. One subject was a subject with authorized access and a basic knowledge of the system. The second subject acted as the attacker and had varying levels of understanding about the functionality of the system. The six study scenarios and the aids used are described in Tables: 4.3 and 4.2.

4.4.1.2. Data Acquisition

Data acquisition involved a broad range of participants which were chosen to ensure a wide range of different physical characteristics were covered. These characteristics were related to height, weight, body shape and two body mass index (BMI) classes 'lean' and 'normal'. Height formed three groups: small, medium and tall. This resulted in a total of six different body types. The test subjects were in¬formed of the procedure beforehand to limit the effects of their behavior on the evaluation process and measurement. A laboratory employee guided them through the process. The participants were instructed to position themselves in the middle of the mantrap portal and turn towards the exit.

While standing, they were told to focus on a point that had been marked on one side of the portal's wall. The laboratory employee started the recording process with the permission of the test participant. Data collection (enrollment, genuine verif./ident. and imposter verif./ident) took place on different days to simulate a real-life scenario. Twelve subjects were enrolled in the data acquisi¬tion process. For the genuine data collection, each test subject was recorded both with and without items (see Table: 4.1). In the case of the attack scenarios (imposter verification), the subjects were divided into groups of 6 subjects where each group alternately played the role of an "authorized person" and an "attacker" (see Table 4.2 and Figure: 4.8).

The mantrap portal allowed for no external distractions in form of sound or light. The attack schemes were shot with two subjects entering the mantrap portal. They had to carry out 6 different scenarios (as depicted in Table 4.2). A total of 21 RGB images over a period of approx. 3-4 seconds were collected during each recording. See Figure 4.14 for examples of attacks in the attack shemes and 4.13 for some examples of genuine verification attempts with- and without objects.

Table 4.7. – Quantities of acquired image sets

Dataset	Recordings	Description
Enrollment	12	12 subjects (2x all body types)
Genuine verification/identif.	48	12 subjects x 4 scenarios (3 with objects, 1 without)
Imposter verification/identif.	216	6 subjects (all body types) x 6 subjects x 6 attack scenarios

4.4.2. The Proposed Methods

The goal was to distinguish if there was more than one person within a mantrap portal and, if so, deny access. In order to guarantee access into a secured area, only one employee can pass through at any one time. In the case

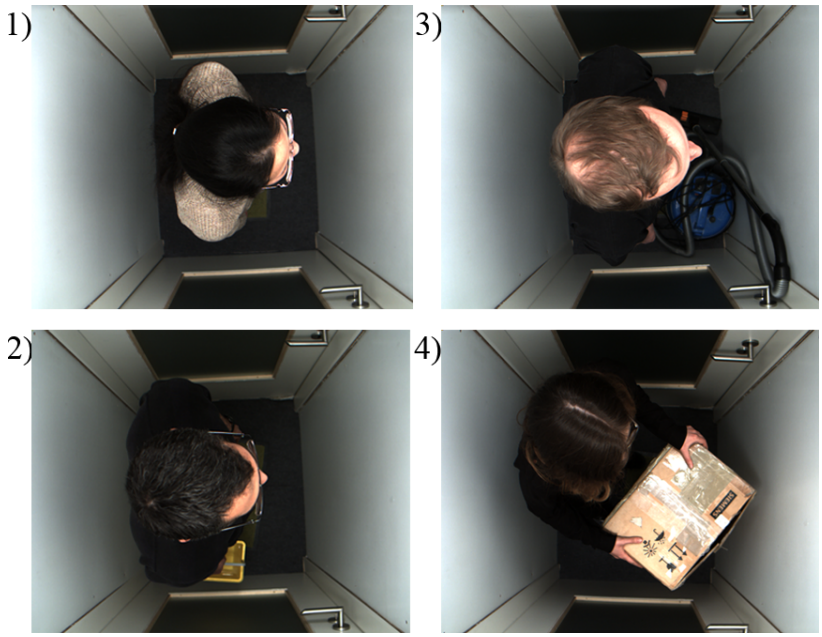


Figure 4.13. – Genuine Verification Attempts with- and without Objects. (1) No Object (2) Plastic Bucket (3) Vacuum Cleaner (4) Parcel

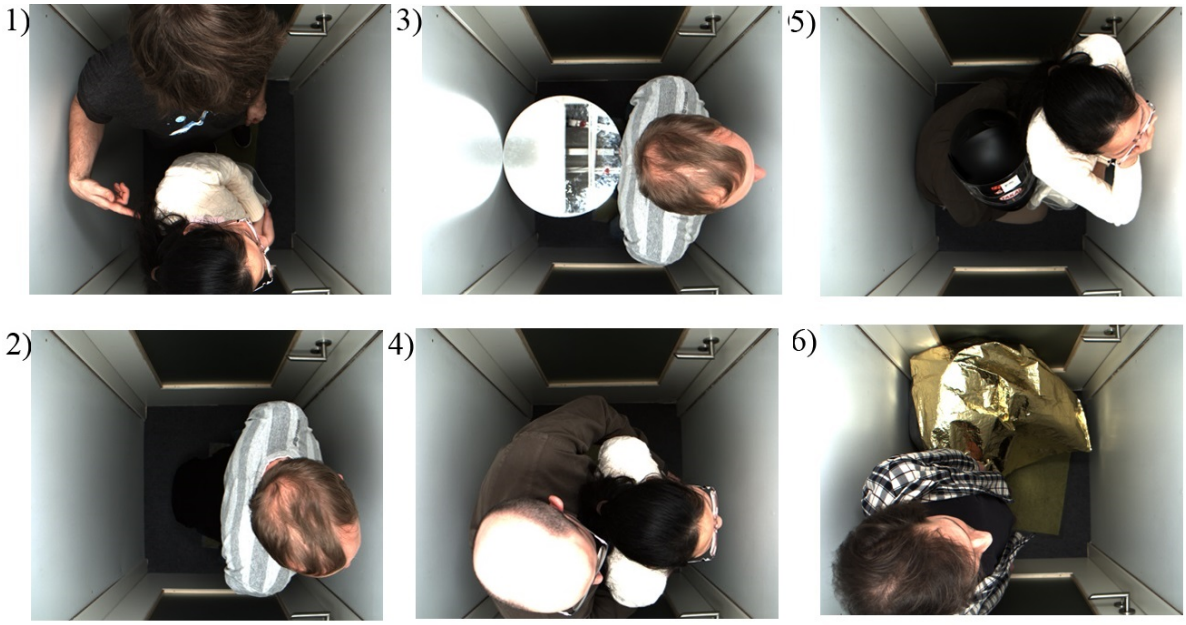


Figure 4.14. – Imposter Verification attempts with different Attackers. The Numeration corresponds to the Attack scenarios shown in Table 4.2.

of attacks where employees are threatened and used to gain access for someone else, access to the secure area should not be granted.

The approach we present in this paper is based on the optical flow method used on RGB color images taken using a monocular camera. The size of the images is 1280 x 1024 pixels. The advantage of using optical flow is its ability to detect micro movements within the images over time. We assume that two persons cannot keep perfectly still within the mantrap even if they try to stay as still as possible.

Micro movements can still be detected using optical flow in images. The difference between one person, one person with an object and two people should be visible in the optical flow collected by the images over time. The exact targeted case stud¬ies will be described in more detail later in this paper.

For each test setting, we collected 21 images over time in the instances $t_0, t_1, ...t_{20}$ and thus were able to determine the optical flow (OF) from time instance to the next time instance denoted as $OF_{t_0:t_1}, OF_{t_1:t_2}, ...OF_{t_{19}:t_{20}}$. For each pair of images we determined the dense optical flow method introduced by Gunnar Farneback in year 2003 [Far03], in which he described a method to determine the gradient field on each pixel within two successive images.

In order to discover the minor movements within the images and to take the spatial distribution into consideration, we fur¬ther divided the whole image I of 1280 x 1024 pixels into smaller image collections of 32 x 32 pixels with the index $idx = 0, 1, ..., 1279$. For each image patch, we calculated two features to be used in the following classification scheme. The first feature to be introduced was the calculated distance within a patch size of 32 x 32 pixels - denoted here as M x N pixels:

$$d_{idx} = \sum_{i \in M; j \in N} d_{i,j} = \sum \sqrt{(x_{i-1} - x_i)^2 + (y_{j-1} - y_j)^2} \tag{4.6}$$

Next we normalize the patches over an image I using:

$$\bar{d}_{idx} = \frac{d_{idx}}{\sum_{idx \in I} d_{idx}} \tag{4.7}$$

We then finally determined the variance of the changes over the summed distance over time where T is the number of time instances:

$$\sigma^2_{idx} = \frac{1}{T-1} \cdot \sum_{t_i=0:20} \left(\bar{d}_{idx,t_i} - \frac{1}{T} \cdot \sum_{t_i=0:20} \bar{d}_{idx,t_i} \right)^2 \tag{4.8}$$

In order to conform to the classifier classification range, we also normalized the variance of summed distances. In such that it is in the range of [0:1]

$$\bar{\sigma}^2_{idx} = \frac{\sigma^2_{idx}}{\sum \sigma^2_{idx}}. \tag{4.9}$$

In this way, no features are preferred due to amplitude. These features are distinct. The micro movement is distinc¬tive because the variance within this specific patch is larger in comparison to the motionless background.

A similar approach was used to extract the features of the variance of angular distribution over the patches. For each patch, the angular variance was determined by using the follow¬ing equation from non-euclidean statistics:

We write each vector of each image pixel in the complex form -as it is easier way to represent vectors- using $z_n = cos(\phi) + i \cdot sin(\phi)$. The mean resultant vector can be given by

$$\bar{\rho} = \frac{1}{N} \cdot \sum z_n \tag{4.10}$$

The length of the resulting vector:

$$\bar{R} = |\bar{\rho}| = \frac{1}{N} \cdot \sum |z_n| \tag{4.11}$$

also gives the variance of the angular distribution over each patch. For strongly ordered vectors, we achieved a length of almost 1 and for strongly disordered vector distribution we got a length near 0. To get the variation over time, we then calculated the variance of each patch over time using equation 4.8:

$$\sigma^2_{idx} = \frac{1}{T-1} \cdot \sum_{t_i=0:20} \left(\bar{R}_{idx,t_i} - \frac{1}{T} \cdot \sum_{t_i=0:20} \bar{R}_{idx,t_i} \right)^2 \tag{4.12}$$

In case of 32 x 32 pixels per patch, we wer able to extract 1280 features over one entire image of 1280 x 1024 pixels to get the variance of the sum of distance and the variance in the angular distribution within each patch. To further consider the changes over time, we determined the variance of these features for a sequence of 21 images taken over time. The overall features space spans a dimension of 2 x 1280 features. One example can be seen in Figure 4.15 where the optical flow from two successive images is depicted. On the right-hand side is an enlarged segment from the entire image, where the micro movement of the head can clearly be seen within the distribution of the flow field. In the training and verification phases, we used either one set of features or fused the two sets using a weighted sum combination to improve the results.

The equal error rate for both sets of features was calculated and used to fuse the scores individually. In the weighted-sum score-level fusion rule, the inverse of equal error rate was used as weight to the scores (EERW)[DO14].

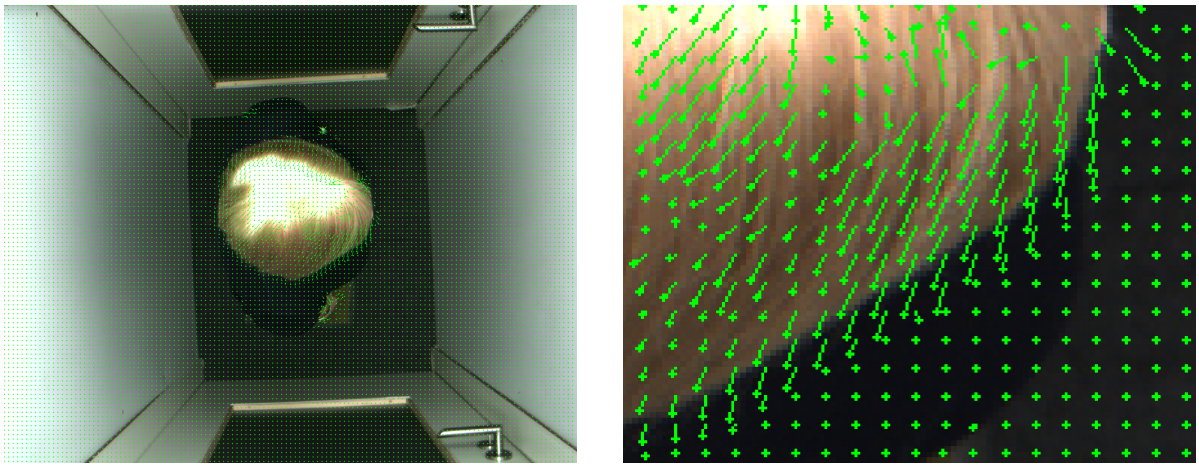


Figure 4.15. – The image on the left depicts the optical flow from two successive image sequences, while the image on the right is a zoomed in section of the left image. The disturbance in the optical flow is caused by the micro movements of the head.

The REAL boosting method was chosen for AdaBoost classification which utilizes confidence-rated predictions and was expected to work well with the categorical feature vectors.

4.4.3. Experiments and Results

4.4.4. Experiments

We tested our novel approach on the acquired database and compared it to following methods used in our previous work. The image data in the other methods was acquired at the same time and corresponds to data set used here.

4.4.4.1. Thermal Imaging [SHK16]

The camera used in this approach offers an adaptive adjust¬ment in which changes of temperature in the camera field of view are highlighted using false colors. In HSV color range, the number of pixels higher than a particular threshold were counted and averaged over all frames. During the period of the recording (120 frames), the variance is calculated and added to the pixel average. Color value and variance were used in binary classification using a distance metric.

4.4.4.2. RGB-D [SWB16]

With this method, change detection and blob detection for RGB¬D images were used to define a shape-model of the single subject first detected in the mantrap. A background model was calculated from the remaining parts of the image and used to extract color histogram features. The Adaboost machine-learning technique was used for classifying the model in the verification without identity claim scenarios.

4.4.5. Results

Firstly, the two proposed optical flow methods: 'dis¬tance calculated' and 'variance of angular distribution', were calculated using patches of the images and processed individually. The evaluation using leave-one-out, 4-fold cross validation showed that the first method performed slightly better. An EER of 6.37% was achieved for the first feature vector and 10.29% for the second one.

The scores of both feature vectors were combined using the 'weighted-sum score-level fusion rule' and the in¬verse of EER as weight. The overall, fused EER of 5.17% was calculated using the same leave-one-out methodology which outperforms all other methods that have been evaluated. The result indicates that the two different feature vectors contain complementary information. Compared to the competitive methods, 'Thermal Imaging' and 'RGB-D', the attack scenarios without additional aids, resulted in relatively high error rates (see Table: 4.8) of around 5%. In contrast, the two scenarios in which a mirror and a sheet of aluminum were used by the attackers, resulted in al¬most a 100% correct detection rate. We assume that the two aids show a lot of movements which results in a high detection accuracy.

The objects carried by authorized single subjects had an impact on the system. We observed that in scenarios where subjects where holding objects in their hands (e.g. a vacuum cleaner) lower accuracy was achieved. There was no measurable correlation between the body size or stature of the attacker and their chance of success. The ROC curve (see Figure: 4.16) shows the archived FPR (False Positive Rate) in relation to the TPR (True Positive Rate) of all verification attempts and in relation to the other methods.mentioned.

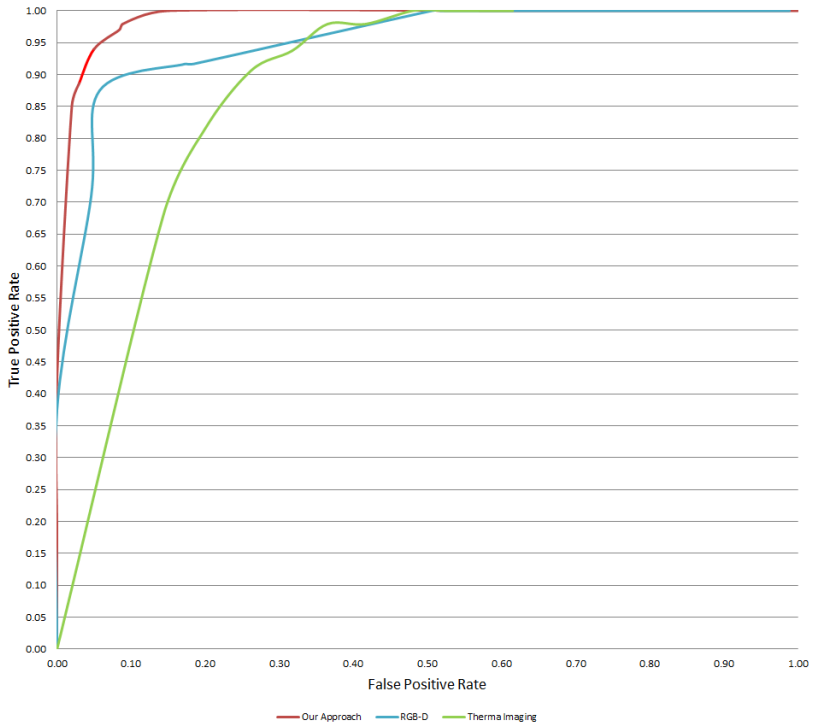


Figure 4.16. – Overall results of the approach presented in this study and competitive methods.

4.5. Feet Verification using Capacitive Sensing

At many every day places, the ability to be reliably able to determine how many individuals are within an automated access control area, is of great importance. Especially in high-security areas such as banks and at country borders, access systems like mantraps or drop-arm turnstiles serve this purpose. These automated systems are designed to ensure that only one person can pass through a particular transit area at a time. State of the art systems use camera systems mounted in the ceiling to detect people sneaking in behind authorized individuals to pass through the transit space (tailgating attacks). Our novel method is inspired by recently achieved results in capacitive in-door-localization. Instead of estimating the position of humans, the pervasive capacitance of feet in the transit space is measured to detect tailgating attacks. We explore suitable sensing techniques and sensor-grid layout to be used for that application. In contrast to existing work, we use machine learning techniques for classification of the sensor's feature vector. The performance is evaluated on hardware-level, by defining its physical effectiveness. Tests with simulated attacks show its performance in comparison with competitive camera-image methods. Our method provides verification of tailgating attacks with an equal-error-rate of 3.5%, which outperforms other methods.

4.5.1. Capacitive Sensing Grid

As discussed in the previous chapters, imaging methods have shown good results on detecting tailgating using top-view mounted cameras, but lack efficiency in several other attack scenarios (e.g. hiding on the floor).

Table 4.8. – EER by Attack Scenario

Scenario	Optical flow	RGB-D	Thermal
All	5.17%	11%	22%
1 No knowledge	4.27%	11%	6%
2 Basic knowledge	4.98%	12%	6%
3 Adv. knowledge	5.10%	22%	6%
4 Helmet	5.12%	7%	26%
5 Mirror	0.01%	7%	56%
6 Sheet of Aluminium	0.01%	12%	28%

Furthermore, they are not applicable to locations with high flow-rates, where fast verification is required. We incorporate a novel approach using a grid of capacitive sensors for detecting and classifying capacitive resistance on the floor to recognize tailgating and provide easier access for handicapped people. Capacitive sensing has different properties that influence their applicability. Especially the sensibility to humidity and distance are challenging. In the following sections, we show how we face this limitation. It contains 1. a short review about sensing theory (see Section 4.5.1.1) 2. our sensor-grid hardware (see Section 4.5.1.2) and 3. our method for classification of the collected data (see Section 4.5.2).

4.5.1.1. Capacitive Sensing Theory

Capacitive sensors are proximity sensors that detect nearby conductive objects by creating an electric field [Bax96]. The technology is based on the capacitive coupling that takes the capacitance produced by the human body to an electrode as an input. This way, it is possible to detect and measure anything that is conductive or has a dielectric difference from air. The measured capacitance is a function of the distance (d) of the object to the electrode, the area of capacitive plates (A), and dielectric the constant (ε_r) of the material between object and electrode; Therefore:

$$C = \frac{A}{d} \cdot \varepsilon_0 \cdot \varepsilon_r \tag{4.13}$$

Capacitive sensing is divided into three categories based on their modes of operation (shunt mode, loading mode, transmit mode). We use loading mode sensors [Smi96], because they have a large range and are easy and cheap to implement. Loading mode sensors deliver a constant current to the attached measurement electrode. The time needed to charge the electrode up to 80% and discharge the electrode down to 20% is measured. If an object approaches the electrode, the capacitance becomes larger and the time needed to charge/discharge the electrode rises.

4.5.1.2. Our Sensor-Grid Hardware

We propose the use of loading mode sensors because they provide a continuous signal for analysis compared to others. The sensor consists of a microchip providing UART (Universal Asynchronous Receiver Transmitter) for communication. A MSP430 micro-controller is used to process the sensor values to binary format. A sensor requires between 1 to 2 mA at 5V, therefore a 5V USB connector is sufficient as a power supply. The hardware shown in Figure 4.17 is the hardware prototype we used. All sensors are connected in a chain to the UART

data-bus which is read at baud rate of 115,200 per second. As the form of the electrode has a high impact on the distance and sensitivity, we have conducted several tests in order to find the right shape (see Section 4.5.3.1).

Figure 4.17. – Used capacitive sensor with UART interface.

We assumed having a square area of 800x800mm as the area to be analyzed. We propose using sensors mounted in the floor of the transit area, located in a grid used for the alignment. The sensors are mounted in the middle of each cell at a distance of 100mm between each sensor. In Figure 4.18 our hardware is shown as a wooden prototype. The grid is placed inverted on the wooden board, so that the sensors are facing the ground. External plywood pieces with specific dimensions are attached to support the whole structure. It stops the breakage of sensors and problems with the wire cabling. As the sensors are not visible to the subjects, they will have no direct impact on their general functionality. The front plate consists of a medium-density-fiberboard with a thickness of 12mm.

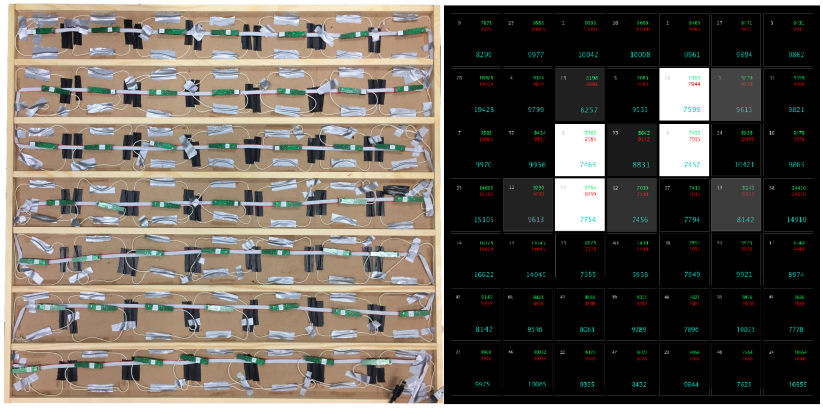


Figure 4.18. – Sensing grid prototype with cable-shape electrodes.

The sensing distance of the capacitive sensors depends on following factors: (1) sensor diameter, (2) sensor design (with/without GND electrode), (3) material of the medium to be detected and (4) the size of the developed body. On one hand, a bigger electrode increases the range of the sensor and reduces the effect of noise in the signal. On the other hand, a higher sensor-range causes indifferences between sensors. Reading sensors at the same time, which are arranged close by, results in wrong capacitance values. The range of sensors used is therefore limited by the density of the sensors on the grid. In our hardware we increased the range of the sensors by reading them in order like a chessboard. Only diagonal neighbored sensors are read at the same time. Consequently the maximal distance between two sensors is reduced to around 140mm. For choosing electrodes, we evaluated the use of a cable ring and a solid copper plate. Our results are shown in Section 4.5.3.1.

4.5.2. Data Analysis

Each sensor receives an edge like signal from the timer that indicates its individual capacitance by counting the number of edges in a defined time window. The timer turns the electrode consecutively to charging and discharging mode (see Figure 4.19).

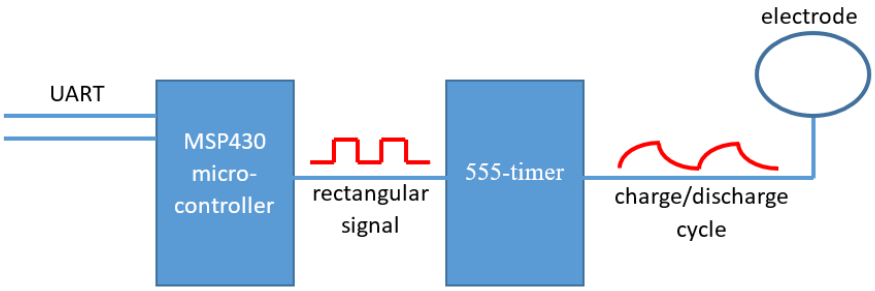


Figure 4.19. – Left: Sensor measuring process and sensor layout Right: Detail of visualization of measured capacitance..

The counted edges in a period of 0.5 seconds get transmitted to the UART bus. Every sensor has its specific ID. All even IDs start measuring the electrode while all uneven ones wait and vice versa. This ensures that sensors that are conducting a measurement do not influence adjacent sensors. The classification software receives a data package, containing single numerical values of each sensor, in intervals of 0.5 seconds. Even, when there are no objects close to the sensing area, the measured values show some differences in delta and amplitude. These are caused by environmental noise and differences in the electronic parts used (e.g. slightly different size of electrode). In order to eliminate the environmental effects, we calculate a baseline value for each sensor. We use the following equation to update the baseline constantly over time:

$$b_n = a \cdot b_{n-1} + (1-a) \cdot x_n \tag{4.14}$$

, where x_n is the current sensor value and b_n is the currently updated baseline value for this sensor. How much the current sensor value influences the baseline will be determined by the factor a.

After subtracting the measured value from the baseline to get the actual signal strength, we normalize the value to a range between 0 and 1. We use a min-max algorithm where min and max are the minimal and maximal measured values over a period of time, ranging from time index $\{i = 1..N\}$ using the following equation:

$$\overline{x_n} = \frac{x_n - b_n}{max(\{x_i - b_i, i = 1..N\}) - min(\{x_i - b_i, i = 1..\})}$$

We propose the use of machine learning for classification of the measured values. We chose SVM with a linear kernel and AdaBoost using REAL boosting as classification methodology. The feature vector, used for training, contains the values of the normalized sensor output. In Section 4.5.3.2 we describe our experiments about the influences of the size of the feature vector on classification accuracy. In these experiments we accumulated the sensor output over time (t) of all sensor to the feature vector receiving $fv = t_0, t_1 .....$ We separated our data into two classes, one subject and more than one subject. In order to detect malfunction of the sensors or transmission, we visualized the sensor output as shown in Figure 4.18.

Based on the different locations where mantrap portals and drop-arm turnstiles are being used, we defined two test scenarios. In case of mantrap portals, subjects can be asked to stand on an exact position which is marked on

the ground. When using drop-arm turnstiles, people usually expect that there is no such position guideline. We therefore acquired data for both scenarios:

1. The access allowed subject is positioned at a marked position
2. The access allowed subject is encouraged to stand freely at a random position

We used a test group of 15 subjects with varying feet-size (between 38 to 48) and body-weight. All recordings are made over a period of 6 seconds (1x49 values per 0.5s), resulting in 12x49 values per recording. One subject acts as an authorized person while the 2nd person acts as the attacker.

We evaluated two attack scenarios:

1. The attacker positions himselves randomly on different challenging positions (on the edge, close to the each other ...)
2. The attacker positions himselves randomly as in scenario 1 and lifts one feet from the ground

The evaluation is performed by training separate classifiers for the two attack scenarios, using Adaboost [VJ01] and SVM classifier. We used the collected feature-vectors of an attack-scenario (two subjects at the same time) and of the different single subjects (see Table 4.9 #1 or #2) for training. We used data collected in 3 seconds (6x49 unsigned integer values) as one-dimensional feature-vector for training and tests.

Table 4.9. – Quantities of Acquired Test-Data.

#	1th Subject at Pos.	2nd Subject at Pos.	Data-Sets
0	marked	none	1080
1	random	none	1800
2	marked	random	1080
3	random	random	1800
4	marked	rand.+ foot lifted	1764
6	random	rand.+ foot lifted	1764

4.5.3. Experiments and Results

We performed tests in order to ensure that our considerations about the chosen hardware are correct. Therefore, we performed laboratory tests about the measuring distance of the capacitive sensors using different electrodes. To ensure that different soles of shoes do not influence the measurement we performed empirical tests which are described in section 4.5.3.1.

4.5.3.1. Sensor Range and Robustness

We evaluated our hardware accordingly to Valtonen at Al. [VMV09] with two different electrodes and with ten persons of different sizes.

We considered the electrodes types: loop and copper plate to be used as electrode. Connecting wires are made as small closed loops to act as an electrode, with one end of the wire connected to the sensor. On the other hand, copper plates in the size of 50x70mm are getting used in comparison. We performed test in range and sensibility using a wet bottle of water as conductive object. We measured the sensor value for different distances and subtracted it by the baseline value. The noise showed for both electrode types, values between 0 and 20 measured over a time period of 2 minutes. Our results (see Table 4.10) show that the sensing range for the copper

plate is higher than for the loop. The measurable distance is of around 100mm which leads us to the assumption to chose a horizontal and vertical distance of 100mm between the sensors as best layout compromise.

Table 4.10. – Measured Capacitance of Electrodes At Different Distances and SNR of Copper Plate.

Object	Distance	Cable Loop (SNR)	Copper Plate (SNR)
bare foot	20mm	7744 (54.83)	1300 (45.26)
bare foot	40mm	10420 (58.02)	4900 (48.01)
bare foot	60mm	11940 (62.16)	7500 (49.98)
bare foot	80mm	12425 (68.22)	8560 (51.65)
bare foot	100mm	12680 (75.35)	8810 (54.23)
bare foot	150mm	12770 (89.12)	9030(59.01)
none	-	13100	9225

In addition, we measured the Signal-To-Noise Ratio (SNR) of the system when the test subject was standing at different distances from the receiver with and without shoes. The SNR depends upon the distance of the feet from the electrode, its type and size. We calculated the SNR using fallowing formula:

$$SNR = 20 \cdot \log \frac{Signal\,Range(SR)}{Noise\,Range(NR)} \tag{4.15}$$

$$SR = max(\{x_i, i = 1..N\}) - min(\{x_i, i = 1..N\}) \tag{4.16}$$

$$NR = max(\{y_i, i = 1..N\}) - min(\{y_i, i = 1..N\}) \tag{4.17}$$

where y_i is the sensor value for a sensor without any feet on it and x_i with feet, over an amount of 200 sensor values (N).

In order to verify that there is no impact of different shoes on the measured values, different kinds of them, like sports shoes, sneakers and woodland shoes are considered. The size of the shoes varied from euro size 38 to 48. The test procedure was done by recording the conductive sensor value of different kinds of shoes individually. We compared those values with the values of the same subjects in bare foot (see Table 4.11).

Table 4.11. – Capacitance of Different Shoes With Copper Plate Electrode.

Type	Material	Sole Height	Sensor Value
sport shoe	plastic	21mm	1993
sneaker	plastic	23mm	2043
woodland	leather	34mm	1845
boots	plastic	38mm	1859
bare foot	-	0mm	2096

We noticed that the SNR decreased somewhat linearly with the distance between the feet and the electrode increases. Comparing the wire electrode with the copper plate, results the copper plate in better SNR then the loop electrode, which stands in contrast to the results of Valtonen at Al. [VMV09].

4.5.3.2. Results

The performance of our method depends on individual user properties, which are: feet position, behavior, feet size and their individual capacitance. We have chosen a quantitative evaluation with a test-group in order to prove our test-setup with realistic data. The evaluation process is carried out by classifying the data into four individual folds and taking 75% (3/4) of the data as training and 25% (1/4) of the data into the test. We calculated the false acceptance rate (FAR), false reject rate (FRR) and EER for all scenarios. We regard the verification case as a closed set, because it only differentiates between 'one subject' and 'more then one subject'.

Table 4.12. – EER of Examined Scenarios.

ID	1th Pos.	2nd Position	AdaBoost	SVM
1	marked	all feet on ground	0%	3.4%
2	random	all feet on ground	0%	5.2%
3	marked	one foot lifted	5.4%	9.1%
4	random	one foot lifted	7.1%	10.3%

Our results show that, the first attack scenario gets always recognized correctly, independently from the position of the access allowed subject (see Table 4.12). Only in the second scenarios, where the attacker lifts one foot, the performance decreased considerably. Classifying the data with linear multi-class SVM, resulted in significantly higher error rates in all attack scenarios.

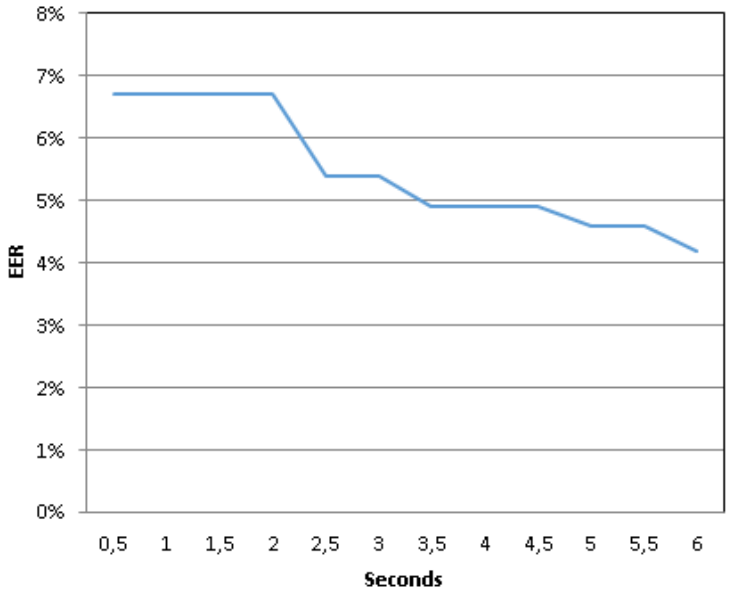


Figure 4.20. – EER in respect to time (one foot lifted).

The time needed for data collection is an important factor of usability. Therefore, we performed tests with a varying measuring period. The length of the feature vector decreased respectively. We started using a feature vector of size 1x49 (1/2 second) and ended using the complete data-set of 1x294 as feature vector. As demon-

strated in Figure 4.20 a change of the error rate of 2.5% is shown within 0.5 (6.7% EER) and 6 seconds (4.2% EER).

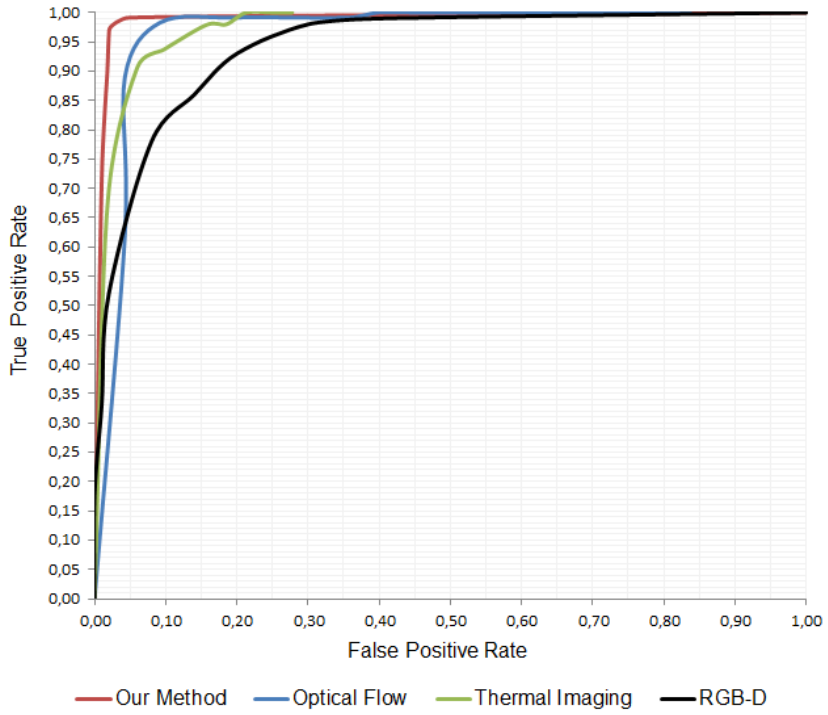


Figure 4.21. – Results in comparison with camera-image methods.

We compared our method with results of other camera-based methods, namely: (1) Optical Flow [SFS*16] (2) Thermal Imaging [SHK16] and (3) RGB-D [SWB16]. As in their work, evaluation was carried out on attack scenarios, where the attacker used additional objects in order to hide himself, only scenarios without any objects were used for comparison. The ROC curve (see Figure 4.21) shows the here presented method, without positioning guideline for the access allowed subject, at random feet position, in comparison with the other approaches. We can conclude that the here presented methods shows a better performance in comparison, nevertheless, might a combination of one of the camera-based methods with our approach be useful in some cases.

4.6. Combining Capacitive Sensing and Imaging

A major risk of an automated high-security entrance control is that an authorized person takes an unauthorized person into the secured area. This practice is called "piggybacking". Known systems try to prevent it by using physical barriers combined with sensory or camera based algorithms. In this paper we present a multi-sensor solution for verifying the number of persons that stand within a defined transit area. We use sensors that are installed in the floor to detect feet as well as camera shots taken from above. We propose an image-based approach that uses change detection to extract motion from a sequence of images and classify it by using a convolutional neural network. Our sensor-based approach shows how user interactions can be used to facilitate

safe separation. Both methods are computationally efficient so they can be used in embedded systems. In the evaluation, we achieved state-of-the-art results for both approaches individually. Merging both methods sustainably prevents piggybacking, at a BPCER of 7.1%, where bona fide presentations are incorrectly classified as presentation attacks.

4.6.1. Methodology

We assume that people have vested interest in coming through the access control point with as little hassle as possible. Previous studies have shown that users "optimize" their behavior in a way that it is convenient to them [PSK*10]. Nevertheless, a system that can be used in practice must find a compromise of usability and safety. But there are other things that need to be considered when developing such a system:

1. The detection method must be flexible with regard to light and clothing as well as different stature of the users.
2. The proposed solution should verify a subject, without the need to claim their identity.
3. Authorized subjects sometimes need to pass carrying different objects, which can be of and kind and appearance.
4. The proposed method should be reasonably fast, in order to be used in an embedded computer close by.

Camera based method do not seem sufficient to guarantee reliable piggybacking detection as they suffer from the limited viewing range of the camera. For this reason, we monitor the floor area by additional verification, verifying that there is nobody hiding on the floor. In the next Section, we therefore present an approach, which interactively controls the that area by means of a capacitive grid of sensors. In Section 4.6.4, we present a new image-based method that combines time-based change detection and convolutional neural networks (CNN) in compliance with the conditions mentioned here.

4.6.2. Dataset

In the image-based method we use the dataset introduce by Siegmund et al. [SFS*16] which includes 60 bona fide verification attempts and 216 piggybacking attacks by 12 different participants (see Figure 4.22). The participants cover a wide range of physical characteristics, like different height, weight, body shape. The attack schemes were shot with two subjects present at a transit area. Six different scenarios are carried out where the attackers showed different approaches to spoof the system and/or hide. Each recording consists of a total of 21 RGB images, recorded over a period of 3-4 seconds. In case of the capacitive sensing grid a test-group, consisting of 12 people with different shoe size (between 37 and 48) was acquired. Each subject was recorded at least once alone and several times with another subject. When evaluating the combined approach, a total of 87 assaults was carried out in different compositions of the test group. The test group was also explained the function of the system and direct feedback of their success was provided.

4.6.3. Capacitive Sensing Grid

In an earlier study, an approach using capacitive active feet detection sensors was presented [SDF*18]. Capacitive sensors are proximity sensors that detect nearby conductive objects by creating an electric field [Bax96].

Since the range of these sensors depend on the size of the electric field, it is possible to detect feet even away from the ground. Therefore, this technology is particularly suitable for the application described (see Figure 4.23). We use the same sensors as the authors of that paper which provide a continuous signal for analysis at a

Figure 4.22. – Some of the Attack Attemps included in the Database.

Figure 4.23. – Schematic Representation of Capacitive Verification.

frequency of 4Hz. In our prototype, we have selected a monitored area of 800x800mm, which acts as the transit area. We mounted $7x7$ sensors in the floor, located in a grid used for the alignment. The sensors are mounted in the middle of each cell at a horizontal distance of 100mm between each sensor. We use a copper plate as electrode because it shows the best ratio of range and sensibility compared to other material. The initial capacitance value of each sensor acts as a baseline value.

4.6.3.1. Active Feet Verification

Previous studies have shown that although capacitive sensors are able to reliably detect feet, but they can not always ensure that they are only a single person. We think that access to high-security areas can be expected to include following interaction of the person entitled to access. First, we propose to ask the to put both feet on the ground. In a next step, the user is asked to lift one foot. So if more than two people are in the area, the intruder would now have to prevent all his feet from touching the ground. We evaluated this procedure in a first scenario by using marked positions on the floor and in a second in which the user is able to freely choose the position of their feet on the floor. Since the measured capacitance changes even when approaching a sensor, we define a threshold ε above which a sensor is considered activated. For this we asked our test group to stand on certain sensors areas without touching the surrounding ones. We then calculated the difference between activated and surrounded sensors for all sensors. We determined ε based on the minimum difference between activated and surrounding sensors plus 20% of the delta. We interpret the sensor grid output as image sg with x rows and y columns. Equation 4.18 applies fixed-level thresholding to the n^{th} single channel matrix $sg^n(x,y)$, $n = 0.0, \ldots, 1.0$ using ε as threshold.

$$dst^n(x,y) = \begin{cases} sg^n(x,y) & if\ sg^n(x,y) > \varepsilon \\ 0 & otherwise \end{cases} \tag{4.18}$$

We get the activated sensors in the resulting image dst where $dst^n(x,y)$ is not 0. In order to detect the lift of a foot, we evaluate for a period of 8 frames whether the previously defined sensors have been activated. If this is the case, the user is asked to lift one leg. Successful validation is achieved when the number of activated sensors has halved in at least three out of eight dst images. We determined the number of only three validation images through experiments that revealed that users need some reaction time. In the second scenario, where users were not given a marked position, successful validation also takes place in two steps. First, the number of activated sensors is counted over 8 frames. The number must not exceed a defined number of sensors. Then it is validated whether the number of activated sensors has halved in the second step.

4.6.4. Image Based Approach

Our method is based on extracting motion features from image sequences using change detection. The reason for this is that a learning algorithm based on the very complex and limited data that we use can be difficult to generalize. Therefore, the complexity must be reduced without losing the information necessary for the detection. Since the background model dynamically adapts to any background, the proposed method is applicable to every background. This is done by finding the difference between the current and previous frames (background model subtraction). For the first three background frames from each scene we create a model of background pixels by K Gaussian's and then check the weight of the mixture representing color proportions. After calculating the background we take the next frames as the foreground and calculate the difference from the background frame.

4.6.5. Motion Detection via Background Subtraction

Our method presented here aims learning models for the cases of bona fide and attack. We assume that the amount, intensity and location of movements in a room differ according to whether one or two people are in it. In doing so, we interpret pixels in the foreground mask as movements. For this reason, objects that are carried by people are not getting included into the feature vector, if they do not move. In our dataset, for each shot situation there are image sequences consisting of 21 pictures collected over 3-4 seconds, we denote them as $i_0, i_1, ...i_{20}$. The first three frames in a sequence are not getting used for feature extraction as they are needed for training the change detection algorithm. For calculating the movement models, we use the frame instances $i_3, i_4, ...i_{20}$ over time. Thereby, we are able to determine the changes detected in image CD from the background model to the next time instance denoted as $CD_{i_3:}, CD_{i_4}, ...CD_{i_{20}}$. We use a Gaussian mixture model based approach [Z*04] where the decision that a pixel belongs to the background is made if:

$$p(x^{\rightarrow(i)}|BG > c_{thr}(= p(x^{\rightarrow(i)}|CD)i(CD)|p(BG)), \tag{4.19}$$

where c_{thr} is a threshold value and the value of a pixel at time i in RGB is denoted by $x^{\rightarrow(t)}$. We will refer $p(x^{\rightarrow}|BG)$ as the background model. The background model is estimated from a training set i_0, i_1, i_2. The estimated model is denoted by $p(x^{\rightarrow}|X, BG)$ and depends on the training set as denoted explicitly. So we calculate the background model using the first three frames and set the learning rate to 0.001 afterward. By doing so, the model is getting updated every 1000 frames which is slow enough to detect all changes background and foreground in the following 18 frames and fast enough to capture changes in e.g. the illumination conditions. We apply change detection frame by frame calculating individual grayscale foreground masks (see Figure 4.24) for

each instance. Another property that we want to depict is the amount of movement. We do so by accumulating the individual foreground masks to a single result image *dst*. Equation 4.20 scales each foreground mask dividing the each pixel of the single-channel mask $CD_i(x,y)$, ranging from $0,\ldots,255$ by 255 and weighting it with a factor of σ. In our experiments we achieved the best results with a σ of 30.

$$dst^n(x,y) = \quad (CD_i{}^n(x,y)/255) * \sigma, \tag{4.20}$$

By doing so, pixel that got recognized as foreground multiple times get a higher value than pixel that have been foreground only for a short time. Therefore, micro movements get visible in the resulting image *dst* (see Figure 4.24).

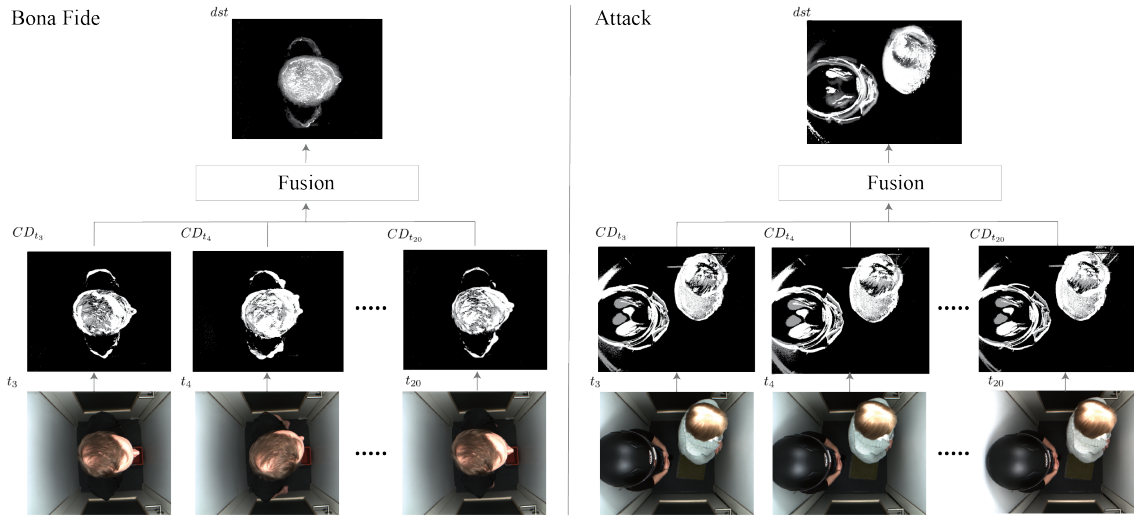


Figure 4.24. – Genaration of Feature Vectors for both Scenarios.

For the classification task we decided to train a convolutional neural network classifier. As we expect that our relatively small dataset could cause the network to show overfitting, we decided to augment the data as follows. First we mirror each *dst* image horizontally and vertically, then we rotate them clockwise 179 times by 2°. On 10% of the data we add additionally Gaussian noise in order to improve generalization of the network. We balance both classes of our trainings data by skipping some rotation steps for the attack image classes. After data augmentation we received 33.124 images of the bona fide class and 41.041 images in the attack class.

4.6.6. Learning a Binary Classifier

Following method is architectural inspired of the proposed Google-LeNet [SLJ*15] but uses an architecture that is quite different from a traditional CNN design like LeNet-5 model. We used inception modules, which perform multiple convolution operations and max pooling in parallel. Therefore its not obligatory to choose a certain convolution kernel size for a certain layer. This approach is not just efficient in classification results but also in computational efficiency. The reason for the computational gain is 1x1 convolution operation which is applied before every 3x3 or 5x5 convolution of the inception module, it results in dimensionality reduction.

Proposed architecture (see Figure 4.25) takes a grayscale image of 256x256 pixel as input. To avoid internal covariate shift batch normalization is used for all convolutional and fully connected layer [IS15]. All weights are

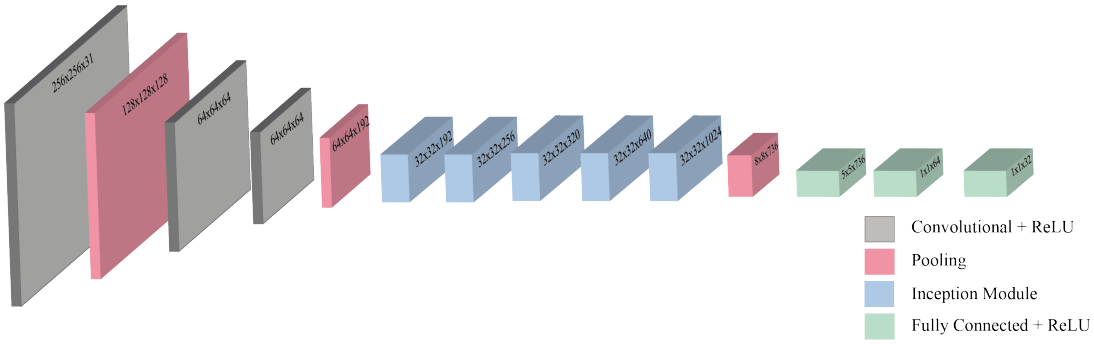


Figure 4.25. – CNN architecture based on GoogLeNet model.

initialized with a normal distribution, using a standard deviation of 0.1 and zero mean. A batch size of 300 is used with the equal representation of both classes. The Training process ran for 200 epoch for each fold of the data set and each epoch contained 3 batches. As both classes are mutually exclusive, a softmax classifier is used with cross-entropy loss function. For all fully-connected layers a dropout of 0.5 is implemented. For stochastic optimization, an Adam optimizer is used with a learning rate of 0.01. Except the final layer, all layers including those inside the inception modules use ReLU activation. Sigmoid activation is used for the final layer.

4.6.7. Experiments and Results

We performed individual experiments in order to ensure that our assumptions about both approaches are correct. The experiments are evaluated based on APCER[1] and BPCER[2]. As the imaging dataset is collected under conditions where the users did not follow any constrains regarding their position, we can not give any separate results. However, it can be assumed that the results on marked position would rather improve here. Due to the small amount of data available, the evaluation is performed using a leave-one-out approach. To make sure that there is no augmented data in the training set, we omitted the complete shot. Using only the imaging approach #4 we achieved an APCER of 1.93% at BPCER of 3.80%. We observed false positives especially in cases where a second person was hiding on the floor. However, it must be said that the attackers had no knowledge about the algorithm used and therefore could not respond specifically. Experimented with the number of sensors activated in the sensing grid we found our, that a number of $2x4$ sensors represent a good compromise between flexibility against great feet and safety. In the case of the marked position on the ground we could not detect any successful attacks in these experiments. However, there were cases in which unintentionally surrounding sensors were activated, which led to an increased BPCER. The approach without marking the position on the floor got circumvented in particular when feet were arranged diagonally to the grid. Since we did not use machine learning in comparison to the comparative study, a single threshold for the activation sensors proved to be too weak. Due to the good results in the case of detection of attacks, we have conducted a fusion on decision level to combine both approaches. Through this procedure, we were able to successfully detect all attacks, but a BPCER of 7.1% must be accepted.

[1] APCER: Proportion of attack presentations using the same PAI species incorrectly classified as bona fide presentations in a specific scenario

[2] BPCER: Proportion of bona fide presentations incorrectly classified as presentation attacks in a specific scenario

Table 4.13. – Results in Comparison with Competitive Methods.

No	Scenario	Marked Pos.		Random Pos.	
		APCER	BPCER	APCER	BPCER
1	Thermal Variance [SHK16]	-	-	20.2%	20.2%
2	Morph. on RGB-D [SWB16]	-	-	11.0%	11.0%
3	Optical Flow [SFS*16]	-	-	5.2%	5.7%
4	BG Substr NN	-	-	2.5%	6.1
5	Capacitive [SDF*18]	5.4%	5.4%	7.1%	7.1%
6	Capacitive LF	0	5.3%	6.3%	18.1%
7	Combining #4 and #6	0%	7.8%	4.3%	15.2%

4.7. Summary

In this chapter, we presented many technical possibilities for how an autonomous mantrap portal could work. The portal used in our tests was described in Section 4.1. In the following Sections 4.2-4.6, we presented a test procedure that shows how we tested whether the methods used make it possible to use the mantrap portal as an autonomous access terminal. The evaluation was carried out on a selection of different people, sexes, and body measurements. Two different knowledge levels were also tested in order to simulate prior knowledge. Tests were also carried out in which the participants were supposed to take objects with them into the portal. We then presented four novel image-based methods and two capacitive sensor-based grids, each trained and used with machine learning. Each of the methods used has shown strengths and weaknesses. The thermal approach 4.2 suffers from a high false positive rate when people take aids into the portal or use them to manipulate the camera. Nevertheless, the evaluation of the presented system shows that thermal imaging is a useful technique in verifying the isolation of people, especially if pose and positioning guidelines (like standing upright, staying still) can be established. False-color RGB-D images are another technique, presented in Section 4.3. The results show very low EER in the case of verification with an identity claim. In addition, the results show that there is no scenario in which an attacker can be sure to overcome the system.

However, there are sensor limitations which reduce its practical usability. The method using optical flow 4.4 performed better than the other two methods. The worst detection rate was also achieved in breach attempts in which the attackers used aids to breach the portal in a targeted manner. The last presented imaging approach 4.6.5 uses background subtraction on an image series and deep learning for classification. A capacitive sensor presented in Section 4.5 tries to detect people sneaking in behind authorized individuals to pass through the transit space (tailgating attacks). Objects brought along, such as suitcases or cleaning equipment, which are occasionally misidentified, cause problems here. In Section 4.6 we showed that a combination of the sensor grid and the imaging method based on deep learning achieved the best results and was able to recognize all piggybacking attacks. A limitation of the system is the requirement of the user to place himself on a position marked on the ground, since the performance otherwise significantly decreases.

5. Novel Classification and Normalization Methods for the Industrial Inspection of Textiles.

Reuse of textiles not only reduces cost, but is also environmentally friendly. However, in the washing and cleaning stages of work cloth and industrial cleaning textiles, manual operation of this task is time intensive and expensive. Automated visual inspection can be a reliable and stable solution in this domain. Similarly, when a defect detection system is used in the manufacturing process, it can ensure a quality product for the consumers. Research on outspread fabrics achieves high accuracy but is not comparable with textiles in voluminous shape. Uneven surface, varying colors, sewing patterns, and weaving of different textile fibers like cotton, linen, and polyester are also some of the challenges in this domain. Automated visual inspection could provide a reliable and stable performance in defect detection. However, its adaptation is still slow, and shows low performance because of several challenges:

1. Similarity of global/local shadow and dark stain defects
2. Need for examination of the large variety of fabric surfaces
3. Variation in the appearance of fiber widths, due to the voluminous shape
4. A defect can take any form and color; the fabric brightness varies
5. Similarity of fiber defects and edge regions

Existing publications tackle this challenge using a decision tree classification process [SKH16][BF15] and pre-processing [SBK16]. The biggest problem that all algorithms face is achieving a high level of generalizability while still maintaining a high recognition rate. Moreover, to ensure real-time detection, it is necessary that the processing steps are as fast and simple as possible. For this reason, the first research questions in this field of study arise.

> Research Question 4: Is an algorithm able to generalize enough to detect defects on different materials?

In order to be able to use an automated, computer-vision-based approach in practice, the following research question also arises:

> Research Question 5: Which system setup in combination with which algorithms allows an evaluation in a comparable time compared to humans?

I present six methods that can therefore be seen as a baseline for fiber defect recognition on uniform textured textiles in voluminous, pile-like shapes. Four woven cotton textile types, each with different fabric, were used as experimental objects. The same cloths were used throughout the presented studies. Defects occur due to heavy utilization and the washing/drying process itself. This includes: stains, bonding, silicon relics, holes, enclosures, dropped stitches, press-offs, or others.

In the first contribution in Section 5.2 the database and its creation process are described in detail in this work. The applicability of the two visual descriptors LBP and SURF in combination with the common classifiers SVM

(Support Vector Machine) and Adaboost is examined. The question to be answered: How will the chosen visual 2D descriptors and classifiers perform on these textiles in a verification scenario? In Section 5.3, a stereoscopic normalization approach is presented. As we incorporate processing of CMOS sensor images with depth channel information, shades and fold can be excluded from that texture and reduce complexity of the images. Section 5.4 presents three novel approaches that recognize and localize fiber defects despite the changing voluminous shape of the textiles and compares them both with each other and with conventional methods. The first is a novel unsupervised approach based on clustering SURF keypoints using an evolutionary algorithm. The second approach combines SURF keypoints with a neural network classification. The third approach (see Section 6) presents a deep learning methodology that is trained using inception modules and patches extracted by a sliding window. We implemented and evaluated these methods also in comparison with other conventional methods, as described in Section 5.4.7.

5.1. Textile Pile Database

In the following sections I will present a novel image database of cleaning textiles captured in a pile. The availability of high-quality and diverse datasets is crucial for the development and evaluation of computer vision algorithms in textile analysis and related fields. This section aims to introduce the creation and characteristics of the proposed image database, highlighting its significance for advancing research in textile analysis and related applications.

5.1.1. Image Acquisition

As there didn't exist a database that fits the focus of this research, the novel database "Textile Pile Database" was created that contains textiles captured in pile like shape. The examined textiles were in a clean and dry condition when they were recorded. Their condition varies a lot because some of them are worn out. Most textiles show therefore dirt, holes or others defects. Furthermore, every textile has one out of three different colors. To classify the test objects based on their different characteristics the ground truth was manually determined by visual inspection. Table 5.1 gives an overview of the different types of textile fabric investigated (see Figure 5.2).

Type	Yarn Density	Fineness	Manufacturing Method
1	9,5x6,5 fibers/cm	50/2 Nm x 5,2/2 Nm	Weaving
2	9,5x12,5 fibers/cm	50/2 Nm x 13,5/2 Nm	Weaving
3	18x26 fibers/cm	80/2 Nm x 250/2 Nm	Weaving
4	26x16 fibers/cm	200/2 Nm x 80/2 Nm	Weaving

Table 5.1. – Qualities of examined textiles.

To protect the laboratory image acquisition process from light entering from the outside, a black box was used. On the inner side of the black box, black molton was attached to protect the fabric from reflection of the box. A uniform sheet of green foam rubber got used as an underlay in order to simplify separation of foreground and background in the segmentation process. For a homogeneous illumination a LED-Ring light with 1950lux was used.

For the two versions of the database, two different cameras were used. In the version 1 of the database for image recording a Camera Nikon D - 90 with a 35mm lens Nikon DX AF - S NIKKOR 1:1.8G was used. In the version 2 of the database we used a soft-box with homogeneous illumination in the image acquisition,

Figure 5.1. – Images captured in Version 2 of TPD (a) Textile after first washing (b) Textile with stains.

Table 5.2. – Camera parameters.

Parameter	Camera used in V1	Camera used in V2 and V3
Resolution	4288x2848 pixels	1280 x 1003 pixels
Focal Length	35mm	8mm
Sensibility	ISO 200	ISO 200
Aperture	F8	F4
Exposure	1/200s	1/100s

to guarantee a controlled capturing process with even lighting. Two synchronized CMOS color cameras with a CMOS 1/1.8" sensor and a resolution of 1280 x 1003 pixels are used for image acquisition. The database contains 910 images of 258 different textiles with and without fiber defects (see Table: 5.4). Fiber-defects are defects that can originate through the manufacturing/furling process (e. g. dropped stitches, press-offs or broken ends) or intensive stress. Most defects of that defect category were caused by intensive use of these textiles in industrial environments and show mostly holes and cut like defects (see Figure 5.8a). Because of diverse uses, the fibers also often show different levels of shading. Shadows caused by folds and overlapping borders show different gradients but have a certain similarity to stains as shown in Figure 5.3a.

In Table 5.2 you can find a detailed description of the parameters used. All test objects were recorded in three different pile-like arrangements. The used data format is JPEG with a low compression rate. The yarn density and fineness is increasing with the type of textile weaving used. This is shown in Figures 5.2a to 5.2d. Furthermore, each type of textile has one of the colors: red, blue, green.

Table 5.3. – Qualities of different types of textiles.

Type	Green	Red	Blue
1	87	3	51
2	69	66	42
3	66	54	24
4	12	3	15

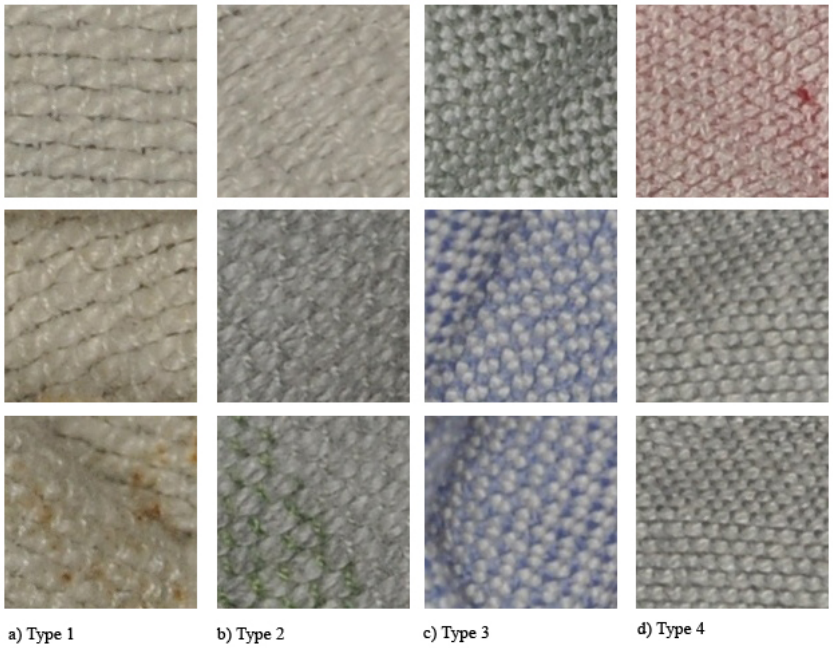


Figure 5.2. – Different kinds of textiles.

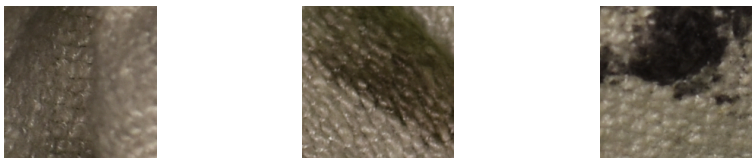


Figure 5.3. – (a) Shadow (b) Absorbed stain (c) On-laying stain (could be e.g. silicone).

5.1.2. Captures in Free Fall

One of the biggest problems with the automatic sorting and quality assurance of deformable textiles is to examine the entire object at once, because the deformability makes the probability of covering individual areas very high. Therefore, and because an orderly placement of the textiles is very time-consuming, I propose the examination in free fall. The here proposed solution is a box where textiles are getting examined in free fall. Camera recordings of the textiles are triggered with the help of the sensors defined below. I used four cameras arranged in the opposite to each other and with light coming from four sides, each lit by lights from the left and right. See Figure 5.4 for examples of images captured in this setup. I recorded a number of textiles (see Table 5.4) using the so called Classifly box (see Figure 5.5). The type of cameras is the same as in Table 5.2 used. Each textile has been recorded three times by all cameras.

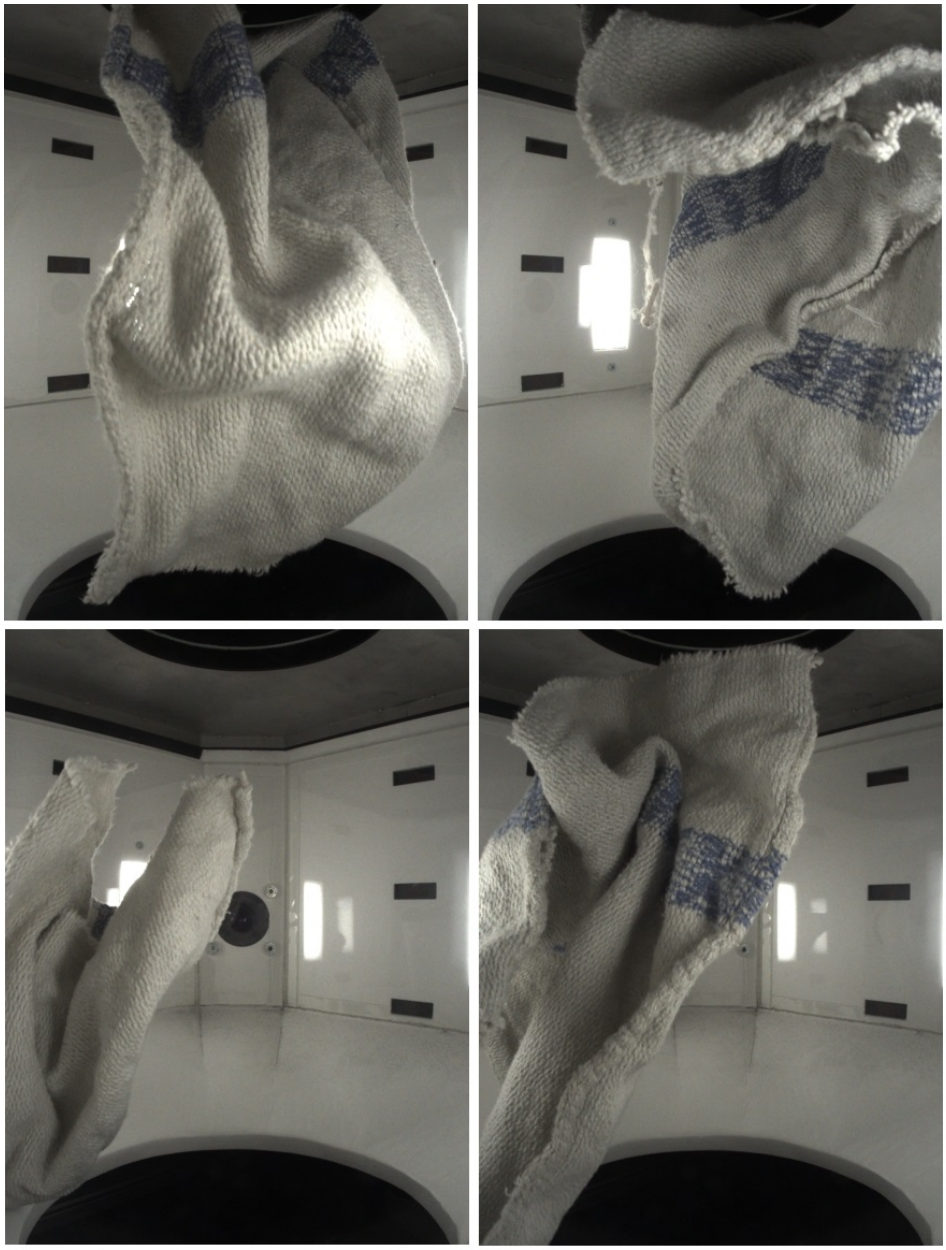


Figure 5.4. – Examples for images collected in V3 of TPD.

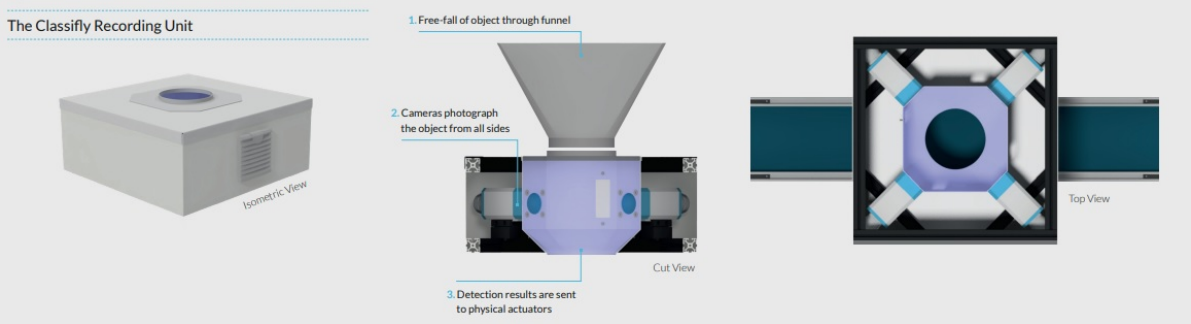


Figure 5.5. – Free fall system "Classifly" used to capture textiles in free fall.

5.1.3. Sensor-Grid

Since the textiles are captured in free fall, where they change shape, multiple images are taken from each camera to minimize the likelihood that certain areas of the textiles are obscured on all images and thus cannot be examined for defects. It is expected that a higher number of recordings examined increases the accuracy. The falling speed of deformable textiles is variable due to the continuously changing shape of the object. Therefore, we propose to use a set of sensors with a logic described below to trigger the camera shots in time. The sensors are positioned on a vertical straight line along the fall axis of the objects, so that the timing of the camera shots is determined by the deflection of one or more sensors to pick up specific parts of the object (and not an empty case if the cloth was too fast).

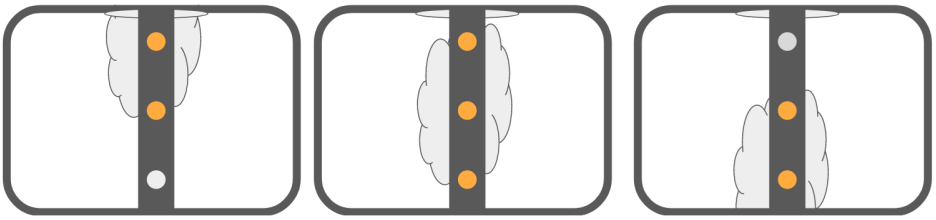


Figure 5.6. – Proposed arrangement of sensors.

Figure 5.6 shows how the sensors are arranged. In the proposed setting an arrangement of three sensors is used. Three recordings are captured, first when the first and second, second when all three and the thirs, when the second and third sensors are activated. By this, regardless of the falling speed of the object, comparable recordings of the objects can be made. As sensory technique, photoelectric barriers using for example infrared or (laser-) light of other wavelength would fit. Also regular cameras, ultrasonic sensors or mechanical switched could be used.

5.1.4. Annotation

The database is annotated in a way that all fiber-like defects (see Figure: 5.8a) are grouped into the class "fiber defect" and all stain like areas on the surface are annotated as "stain" (see Figure: 5.8b). Images of textiles that show stain and fiber-like defects (like holes) are annotated as stain and fiber. There are three versions of the textile-pile database. The main difference, beside the different capturing environment between version 1 and

version 2 are the images with disparity map. In version 2 for each image pair a disparity map is calculated. The process and all details about how we calculated it can be found in 5.3.1.1. In version 3 each textile has fallen 4 times through the box and was recorded 4 times, one time by each camera. The quantities of images and textiles for all different classes of all versions of the database you can find in Table 5.4.

Table 5.4. – Quantities of Images and Textiles in Version 1-3 of the Database.

Database	Defect	Images	Depth Maps	Textiles
Version 1	Stain	113	0	166
Version 1	Fiber	114	0	173
Version 1	Stain and Fiber	112	0	171
Version 1	None	97	0	27
Version 2	Stain	302	156	112
Version 2	Fiber	310	155	98
Version 2	Stain and Fiber	300	150	88
Version 2	None	300	150	72
Version 3	Stain	1.792	0	112
Version 3	Fiber	1.568	0	98
Version 3	Stain and Fiber	1.408	0	88
Version 3	None	1.152	0	72

These textile patterns contained different characteristics such as holes and stains and texture properties like: shadows, different kinds of edges, cuts, open ends, folds etc.. The examined textiles were in clean and dry, but used condition, therefore fibers show different levels of brightness (see Figure 5.1). Patches belonging to the background were rejected using their entropy value. Every patch was labeled manually by assigning them to the class 'stain' or 'other'. Images were captured from a top-view perspective, therefore, some stains might be hidden (e.g. if they are in a fold or on the bottom side). Iterations in which the textile is physically moved into a different position and then classified again could solve this problem.

5.2. Textile Defect Detection via Handcrafted features

Known systems require the fabric to be flat and spread-out on 2D surfaces in order for it to be classified. Unlike other systems, the presented method was examined to classify textiles when they are presented in piles and in assembly-line like environments. Technical approaches have been selected under the aspects of speed and accuracy using 2D camera image data. A patch-based solution was chosen using an entropy-based pre-selection of small image patches. Interest points as well as texture descriptors combined with principle component analysis were part of this evaluation. The results showed that a classification of image patches resulted in less computational cost but reduced accuracy by 3.67%.

5.2.1. Approach

In the following subsections, I present methods employed for textile defect detection. The focus is on developing an effective system to identify and classify defects in textile fabrics. This section provides a comprehensive overview of the system architecture and the methods utilized for pre-processing and defect detection. Fur-

thermore, a method using filtering oo patches with lower quality and information content for classification is presented.

5.2.1.1. System Overview

The evaluation presented in this work consists of: segmentation, patch extraction, pre-selection, feature extraction, classification and fusion. The individual steps of the process are shown as a pipeline in Figure 5.7. The system has been evaluated including and excluding the steps: patch extraction and pre-selection. For feature extraction the local interest point descriptor SURF, as well as the LBP (local binary pattern) descriptor were used. In the classification process the classifiers Multiclass SVM and Adaboost were evaluated. When using patches instead of the full image, these patches got preselected using the Shannon entropy value. The results of the classification are fused in the Decision-Level-Fusion step.

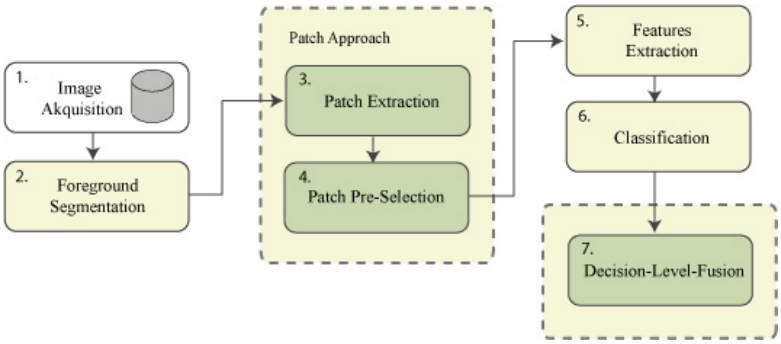
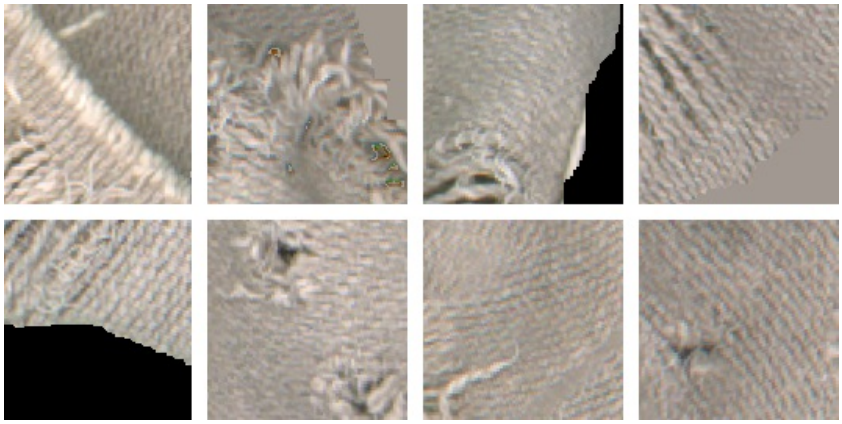


Figure 5.7. – Program flow diagram.

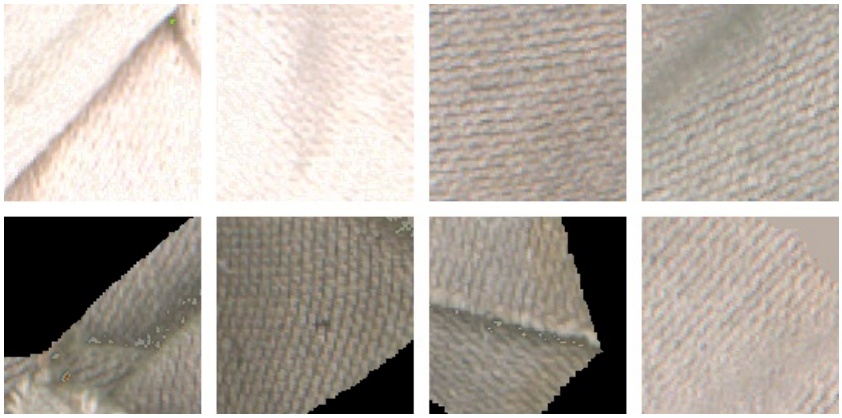
5.2.1.2. Description of used Methods

For visual distinction of the cloth quality between the types listed in Table 5.1 two properties could be used throughout. First, there are marker strips available, differing in every cloth quality; secondly, the weaving of each textile is arranged differently. The cloth fabric can be visually differentiated by its fineness, yarn density and its total mixture. When there are pile-like, uncontrolled arrangement of textiles on the assembly line, a color bar is not always visible and therefore the type of fabric is used as a feature for classification. However, the evaluation of this property requires a high recording quality and a correspondingly high resolution. For this reason screen tests were conducted to determine the minimal resolution with enough features to distinguish the different types of fabric. The Version 1of the TPD is used for this experiment. The images have been divided into patches (see Figure 5.8) and was examined by humans on their distinctness. The tests have shown that a resolution of 4288x2848 pixels (aspect ratio of 4:3) within a receiving area of 30x40cm is optimal. Using the analysis of e.g. “texture spectrum” or “interest points" based features these discriminative properties can be evaluated for a selection. To perform a classification based on the texture of the images, different approaches were tried. These can be differentiated by the used image parts, the features used and the classifier. Table 5.5 gives an overview of the analyzed techniques.

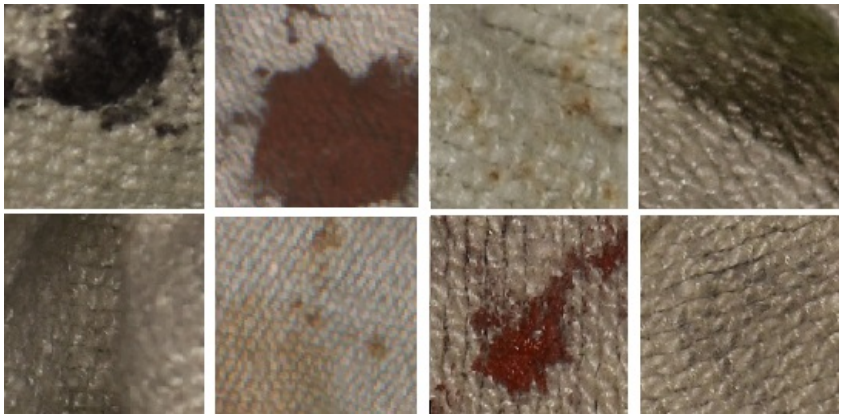
Two different image input data formats were investigated. First the use of the full images in high resolution of 4288x2848 pixels, secondly extracted parts of the image represented in patches of the size 128x128 pixel were used. The idea behind using patches instead of the full image relies on the assumption that pre-selecting image parts which stores more discriminative information than others will provide a more reliable classification.



(a) Patches with Fiber-Defects.



(b) Patches without Defects.



(c) Patches with Stain Defects.

Figure 5.8. – Examples of patches showing different defects.

Image Part	Feature	Classifier
Full Image	SURF (Bag of Words)	SVM or Adaboost
Full Image	LBP and PCA	SVM & Adaboost
Patches	SURF (Bag of Words)	SVM or Adaboost
Patches	LBP and PCA	SVM or Adaboost

Table 5.5. – Different approaches examined in this work.

5.2.1.3. Preprocessing

In the segmentation step (1) the background is separated using the well-known chroma-keying method. The morphological operators erosion and dilation got used to exclude smaller artifacts in the background from the foreground. In the patch extraction (2) the images are split into patches with the size of 128x128 pixels.

5.2.1.4. Pre-Selection

In this approach, a so-called entropy value is determined which characterizes the image quality. Unlike a training based approach the decision is made using a single value. The pre-selection is therefore being used with a threshold value. The so-called Shannon Entropy Value is a value that measures the information content in data and is usually used as a measure of image quality. Based on that, an approach for patch selection is applied to choose the patch with higher entropy, i.e. higher quality and information content. The entropy of a patch I is calculated here by summing up the entropy of each of the three channels of the image. The entropy of each image channel is the sum of all pixel values probability $p(i)$ multiplied by log_2 of those probabilities. The probability of a pixel value $p(i)$ is obtained by calculating a normalized histogram of the possible pixel values (here, $i = \{1,\ldots,2^8\}$). The entropy of a 3-channel, 8-bit image can be formulated as:

$$E(I) = -\sum_{C=1}^{C=3}\sum_{i=1}^{i=2^8} p(i)\log_2\left(p(i)\right) \tag{5.1}$$

The entropy values of all patches are calculated and a number of patches with the highest data content were selected. In experiments, with the selected image size and clustering, a number of seven patches proved to be the best. It is expected that the patches with the highest entropy value also contain the most relevant information that helps to distinguish e.g. holes and stain while patches with low entropy value often represent monochrome patches with shadows or background.

5.2.1.5. Feature Extraction

In feature extraction, the image is assigned to a class representing a textile fabric (see Figure 5.9). In the approach using patches, all divided image patches describe one single class. The well-known SURF features [BTVG06] have shown their effectiveness in many recent papers dealing with object recognition tasks [YJHN07]. SURF Features are scale-invariant and robust against rotation, translation and changing lightning conditions. Therefore they are applicable for the detection of invariant features. Before feature extraction, images were converted into a gray-scale representation and histogram equalization was applied. Then a set of interest points is extracted using the fast hessian detector. The kind of extracted feature points is specified using a library of 120 images. These features were further processed within a 'bag of words' approach using a 64- dimensional Vector as a descriptor.

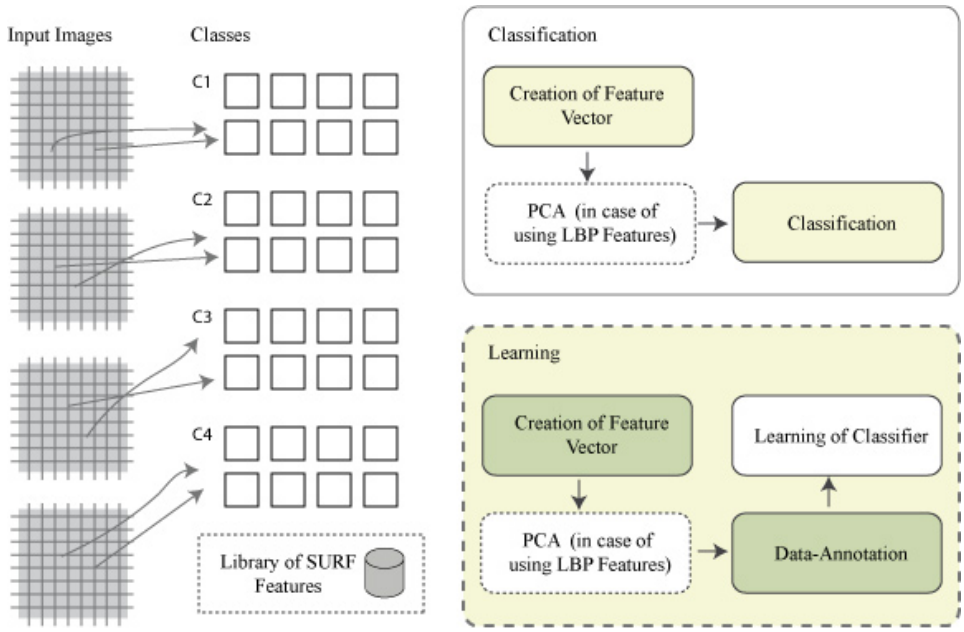


Figure 5.9. – Feature extraction process.

Local Binary Patterns (LBP) are used to analyze texture spectra and is used for classification in computer vision. Its strength is its extreme tolerance towards brightness changes, since only the local gray value changes are considered. They have been applied on different tasks in the field of image recognition [LWZ13] and achieved high detection rates[WHY09]. A darkening of the image (e.g. in case of shadows occurring in folds of the textile) has no more negative influence. After gray-scale conversion and histogram equalization the image is divided into blocks and each pixel of a gray-scale image (except marginal areas) is assigned to a new 8-bit value. This value is calculated from 8 neighboring pixels of the current pixel:

$$LBP = \sum_{n=0}^{7} s(i_n - i_c) \cdot 2^n \tag{5.2}$$

Where p is the subject pixel, i_c is the gray value of the pixel in the gray values of adjacent pixels (continuously starting left above p and then clockwise). On the cells, a histogram is calculated, normalized and interconnected. Eigenvalues, eigenvectors and mean were calculated from a number of histograms using a subset of the database. PCA (Principal Component Analysis) is employed to reduce the dimensionality of the resulting histograms by projecting 300 components. The resulting histograms are used as feature vector for training and prediction.

5.2.1.6. Classification

The image classification in one of the core components of the recognition process. The used machine learning approach requires the training of a classifier. Changes in the previous steps can have a high impact on the recognition results. The previously described features are stored as feature vectors $F : F = f_1, f_2, f_3, \ldots, f_{300}$. For further processing all feature vectors are stored in form of a matrix. Four classes representing the four different textiles are used in the identification scenario. In the verification scenario only two classes are used. One

representing a certain textile type (class 1) and another represents all other classes (class n). For classification of the resulting feature vectors the classifiers Adaboost and Multiclass SVM are used. Five classes cross validation folders is used to verify the classification results. When only patches are used, seven of them were selected from each textile image and their scores fused to a single decision. Verification and identification scenarios have been evaluated.

5.2.2. Results and Experiments

The identification scenario was evaluated with a qualitative selection of image patches (patches), as well as without such a pre-selection (full-image). The pre-selection was thereby done using the Shannon-entropy value. The accuracy indicates the successful differentiation between the 4 classes (True Positive Rate). It was tested against a data set of 537 images. The images were equally distributed over five subsets. For each training of a classifier four subsets were used for training and one for testing. The results in identification show that the approach using patches resulted in a weaker performance compared to the one using the full image. The SURF interest point features show a better performance than LBP features. The reason for this may lie in their scale and rotation invariant characteristic. In the verification scenario the same data set was used as for the identification scenario. As Multiclass SVM outer performed the Adaboost classifier by an average of 3.67% accuracy the verification results are only shown for the Multiclass SVM classifier. The results show clearly better accuracy for all textile types and a difference of only 2.89% accuracy between the patch based approach and the approach using the full image. A possible reason for the poor performance of the approach with pre-selection of pieces of cloth is the kind of information excluded by the algorithm. It can be seen that discriminative information is stored in even patches with lower entropy. The speed of the algorithm using SURF features on image patches on an Intel Core i7 4770 is 503ms. The approach using the full image instead of patches is 923ms.

Image Size	Feature	Classifier	Accuracy
Full Image	LBP/PCA	MC SVM	65.52%
Full Image	SURF	MC SVM	86.43%
Patches	LBP/PCA	MC SVM	59.9%
Patches	SURF	MC SVM	85.41%
Full Image	LBP/PCA	Adaboost	63.96%
Full Image	SURF	Adaboost	82.10%
Patches	LBP/PCA	Adaboost	59.72%
Patches	SURF	Adaboost	80.33%

Table 5.6. – Classification accuracy in identification scenario.

5.3. Textile Defect Detection via Stereo Image Normalization

The visual detection of defects in textiles is an important application in the textile industry. Existing systems require textiles to be spread flat so they appear as 2D surfaces, in order to detect defects. In contrast, we show classification of textiles and textile feature extraction methods, which can be used when textiles are in inhomogeneous, voluminous shape. We present a novel approach on image normalization to be used in stain-defect recognition. The acquired database consist of images of piles of textiles, taken using stereo vision. The results

Image Size	Type	Feature	Accuracy
Full Image	1	LBP/PCA	94.68%
Full Image	2	SURF	89.91%
Full Image	3	LBP/PCA	96.56%
Full Image	4	SURF	94.96%
Patches	1	LBP/PCA	90.01%
Patches	2	SURF	89.26%
Patches	3	LBP/PCA	92.14%
Patches	4	SURF	94.95%

Table 5.7. – Classification accuracy in verification scenario using Multiclass SVM.

show that a simple classifier using normalized images outperforms other approaches using machine learning in classification accuracy.

5.3.1. Normalization Approach using Stereo-Vision

As discussed in the previous sections, competitive methods perform well on ho¬mogenous (spreaded-out) fabric, but lack in case of shadows caused by ambient occlusion, which look similar to stains. We propose a novel approach using a disparity map image for finding and normalizing these areas. It contains 1. the calculation of the disparity-map (see Section 3).1, 2. the recognition of folds (see Section 3.2), 3. the normalization of shadows around folds in these areas (see Section 3.3), 4. the definition of the fiber mean color (see Section 3.4) and 5. the shadow classification and normalization (see Section 3.5).

Figure 5.10. – (a) Disparity map (b) Background Mask.

5.3.1.1. Calculation of Disparity Maps

A disparity map contains depth information from two dimensional images. Stereo geometric images without radial and tangential distortions on the same camera scan-lines with corresponding epipolar lines were used to create a disparity map. The intrinsic and extrinsic parameters were defined by using a calibration pattern and basic geometrical equations. The disparity d is calculated as follows:

$$d = \frac{f \cdot b}{Z} = |x_1 - x_2| \tag{5.3}$$

x_1 und x_2 are points on the image plane corresponding to a scene point in 3D, b is the baseline (distance) of the cameras, Z is the depth of a point and f is the focal length of the cameras. Equation 5.3 shows that disparity

values are inverse proportional to the depth of a point Z. The far points have low disparity and the close points have a high disparity. Furthermore, the disparity is proportional to the baseline b. A larger baseline results in a higher disparity. The image resolution used, allowed sufficient accurate disparity measurements. For the calculation of the disparity map, we use the non-parametric rank transform [Hir05] and semi-global matching [Hir05] was also used. These stereo matching methods outperformed local methods in terms of disparity map quality. We used simple median filtering to remove salt-and-pepper noise from the disparity map images.

5.3.1.2. Fold Recognition

The disparity map and the corresponding rectified image were used in the following process (see Figures 5.1 and 5.10a). We segmented the background using the chroma-keying method on the rectified image for masking. The morphological operators erosion and dilation were used to exclude smaller artifacts in the background from the foreground. The foreground and background separating mask was applied on the disparity image.

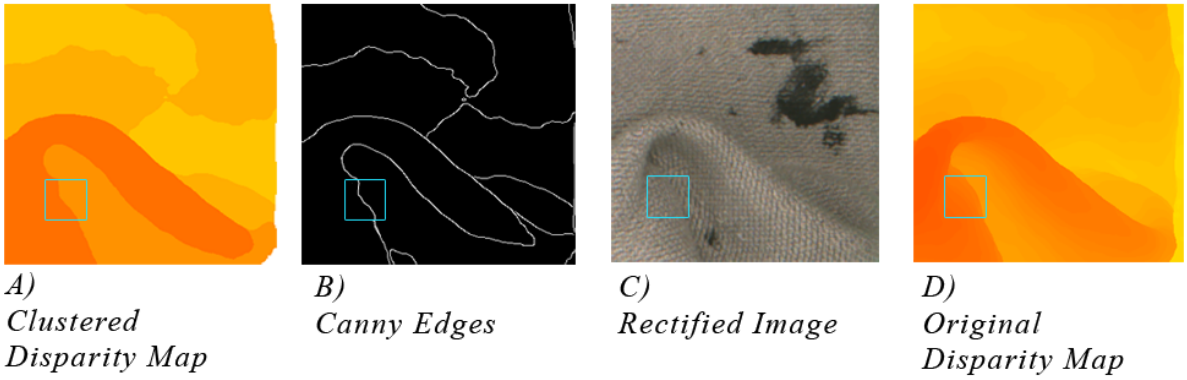


Figure 5.11. – Pipeline showing stain recognition and shadow reduction.

Folds are the only source of shadow in the image but not all folds can be seen as edges on the disparity map image. In deep and wide valley like folds (see Figure 5.11C) of piled fabric inparticular, the shadow size is difficult to estimate. Some shadows also occur on fluent gradients along heights (see Figure 5.11C). Edge detectors like Canny and Sobel are therefore not sufficient to detect shadow reliably. Other approaches discussed in literature either require a 3D mesh for shadow simulation or are too general to be applied on disparity maps. Instead the following method based on color quantization is proposed to detect shadowed folds in disparity maps. K-means clustering was used to reduce the number of colors in the image. Following the approach of Arthur and Vassilvitskii [AV07], centroid values were used for each color channel. These were applied to all pixels of a reshaped image array of Mx3 size (M = # of pixels). After reshaping it back to the shape of the original image, that resulting image had a specified number of colors.

Experiments showed that the transitions of five colors within the defined range of false-colors (and their corresponding disparities) are highlighting edges with folds more accurately than by applying common edge detectors. To identify transitions that do not result in shadows:

a) Their corresponding color in the rectified image was compared to the most common color of the textile (which is assumed to be not a defect) using the invariant color model c1c2c3 [GS99] which has shown good results in shadow identification [SCE01].

b) The difference in altitude within a certain region along a transition was defined and used for thresholding.

5.3.2. Normalization of Shadow around Folds

The clustered disparity map A (see Figure 5.11A) is used to detect edges by applying the Canny edge detector on the color clustered disparity map. Borders in the resulting binary image B (see Figure 5.11B) have a line width of 1 pixel. The outer contour edge of the textile was removed from B by calculating and subtracting the bit-wise conjunction of the inverted background mask (see Figure 5.10b). Image B containing the remaining edges of the disparity map was used as indicator of folds two times:

a) To define the mean color of the fiber by using B in pre-processing.

b) To verify if regions along folds show shadow.

5.3.2.1. Definition of the fiber mean color

For the definition of the mean fiber color representative, areas of the image need to be found and analyzed. Shadow and other defects interfere with the calculation of the mean fiber color. To reduce the negative impact of these, edges in the disparity map were used to exclude these areas extensively. Morphological operator dilation with a rectangle as structuring element and a factor of 12 was used to broaden the line in the binary image (see Figure 5.11*B*). The pixels with positive color values in B were removed in the corresponding rectified image (see Figure 5.11C) by using inverted bit-wise conjunction (bit-wise and) between B and the rectified image. Thus, the resulting image does not contain any edges. Then on that image k-mean color quantization was used to define the color which occured most often δ. It is assumed that the color value δ is the non defect color of the fabric.

5.3.3. Shadow Classification

In order to verify if regions along folds show shadow, image B gets transformed into a vector of coordinates using topological structure analysis. By following the coordinates of the lines stepwise, square bounding boxes representing the region of interest r (see Figure 5.11A-C) with side length of l were created using the line coordinate as the centerpoint. The coordinates of the line within r were stored in a vector v.

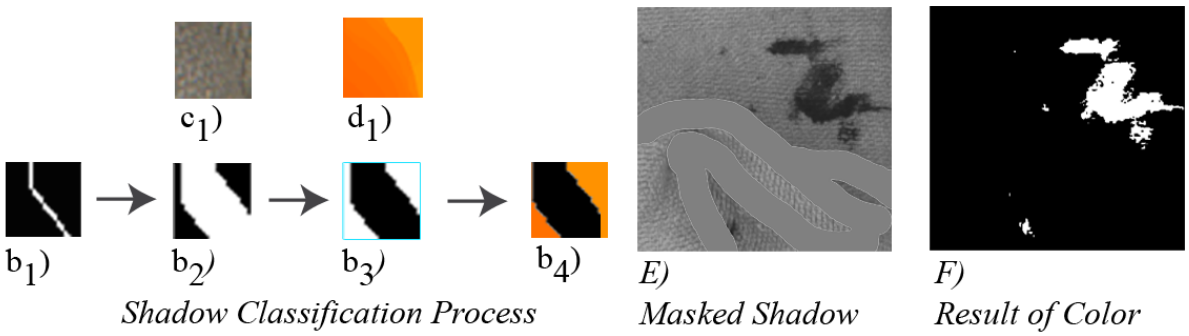


Figure 5.12. – E: Result of Shadow-Normalization, F: Result of Color-Range Threshold, b_1-b_4: Pipeline to verify that a ROI shows a certain amount of Depth and Darkness in order to classify it as Shadow.

The line coordinates v were then used to extract Image c_1 from the rectified image (see Figure 5.12c_1). The color α of the image c_1 was compared with the previously calculated color δ that occurred most often in the image. The deviation between the two colors needed to be higher than a threshold θ in order to classify C_1 as shadow.

$$\alpha < (\delta+\theta) \tag{5.4}$$

The difference in altitude within the region of interest r of the fabric is a second requirement for shadow. The corresponding regions of r were extracted from B into a new matrice b_1 (see Figure 5.12b_1). On this resulting binary image, the morphological operation dilation with a rectangle as a structure element and a factor of 3 was applied (see Figure 5.12b_2). Then the image was inverted bit-wise so that two areas with positive binary values arose (illustrated in white color shwon in Figure 5.12b_3). These two areas represent both sides across a fold. The coordinates of the two areas of binary value 1 were now individually stored as vectors a_1 and a_2. The inverted edge mask image b_3 was then applied via the 'bitwise-and' operation on the corresponding region in the unclustered disparity map (see Figure 5.12d_1). The difference between the mean color value of a_1 and a_2 η (using the H channel of the HSV color space) represent the difference in altitude between both sides across the edge (see Figure 5.12b_4).

$$\|\sum_i a_1 - \sum_i a_2\| < \eta \tag{5.5}$$

Experiments showed that η higher than 10 is adequate. In order to threshold stain defects in the rectified image, image C was converted to gray-scale. The morphological operator dilation with a rectangle as structuring element and a factor of 12 were used to stretch the positive values in the binary image B. Positive values were then replaced by the mean color α and used as mask on image C where pixel color values of B are $\neq 0$ (see Figure 5.12E).

5.3.4. Experiments

We tested our novel approach on the textile pile database V2 of woven cotton textiles and compared the results to state of the art methods on our database.

5.3.4.1. Feature Extraction

Following features were used in the evaluation of the described application:

5.3.4.2. LBP

Local binary Patterns (LBP) features have shown their performance in previous work dealing with textile fiber classification tasks [SKH16]. The used LBP type [ZP07] is invariant against rotation and gray-scale and shows a relationship between pixel and its surrounding. It fulfills the requirements in aspects of computational costs compared to other scale-invariant LBP variants[LLYY12]. Experiments on a subset of the database showed that a radius of 3 and a block size of 32 pixels is ideal for the used database. A histogram of rotation-invariant binary patterns for blocks of 32 pixels was calculated and concatenated to a feature vector. To reduce the dimensionality of the feature vector, Principal Component Analysis (PCA) was applied to a subset of the data set. Experiments have shown that reducing the data set to 300 components gives the best results.

Color Histogram A color histogram represents the distribution of colors in an image. The optimal discretization of the colors into bins was calculated iteratively and evaluated with 4-fold cross validation. A histogram was calculated on the HSV color model for every channel and then concatenated with each other. The evaluation showed that the usage of 130 bins gives the best results for the data-set used. **LBP + Color Histogram** The histogram was calculated using 130 bins for each HSV channel, concatenated with each other and the LBP features vector.

SIFT/SURF Bag of Words The local interest point descriptors SIFT and SURF [BTVG06] have shown in many applications their effectiveness as local feature detectors and descriptors for non-rigid 3D objects [ZWWY16]. They are scale-invariant and robust against rotation, translation and changing lighting conditions. A set of interest points was extracted following the Bag of Words (BOW) approach of Siegmund et al.[SKH16].

SIFT/SURF BOW + Color Histogram As local interest points do not include any color information, a color histogram using 130 bins was computed and concatenated to the SIFT/SURF feature vector.

Color Threshold Since stains defects on the given textiles contain other color than the fabric mean color, a simple binary threshold approach using a threshold on the gray-scale color value of each pixel was used. The amount of black pixels in each patch classifies them, instead of a machine learning classifier.

5.3.4.3. Classification

In the experiments using machine-learning, SVM and AdaBoost performed best among the classifiers: SVM, Random Forrest, AdaBoost and JRIP. Therefore, SVM with the SMO (Sequential Minimal Optimization) extension and AdaBoost were chosen in the experiments as classifiers. The optimal SVM parameters were defined using cross-validation, choosing the optimal parameters for C, gamma, p, nu, coef0 and degree. The REAL boosting method was chosen for AdaBoost which utilizes confidence-rated predictions and was expected to work well with the categorical feature vectors. All features and classifiers were evaluated with-and without use of the proposed shadow normalization method. In the color threshold approach, a color range around the mean-color: α was specified to define the texture color. Pixels with a gray-scale color value higher or lower than the range where considered to be defects (see Figure 5.12*F*).

Table 5.8. – Equal Error Rates in Stain Defect Recognition.

Methodology	SVM w/o norm.	AdaBoost w/o norm.	Color Th w/o norm.	With norm.
Color Histogram	40.15 %	15.77 %	-	13.54 %
LBP + PCA	39.51 %	30.54 %	-	28.14 %
LBP + PCA + ColorHist.	37.14 %	19.68 %	-	15.05 %
SURF BOW	32.56 %	20.11 %	-	25.32 %
SURF BOW + ColorHist.	40.02 %	16.15 %	-	20.91 %
SIFT BOW	28.56 %	18.35 %	-	20.46 %
SIFT BOW + ColorHist.	35.36 %	15.03 %	-	10.90 %
Full Image	-	-	18.74 %	5.68 %

5.3.5. Results

As shown in Table 5.8, almost random classification results were archived using LBP features with SVM and AdaBoost. This approach showed poor performance when distinguishing between stain characteristics and others

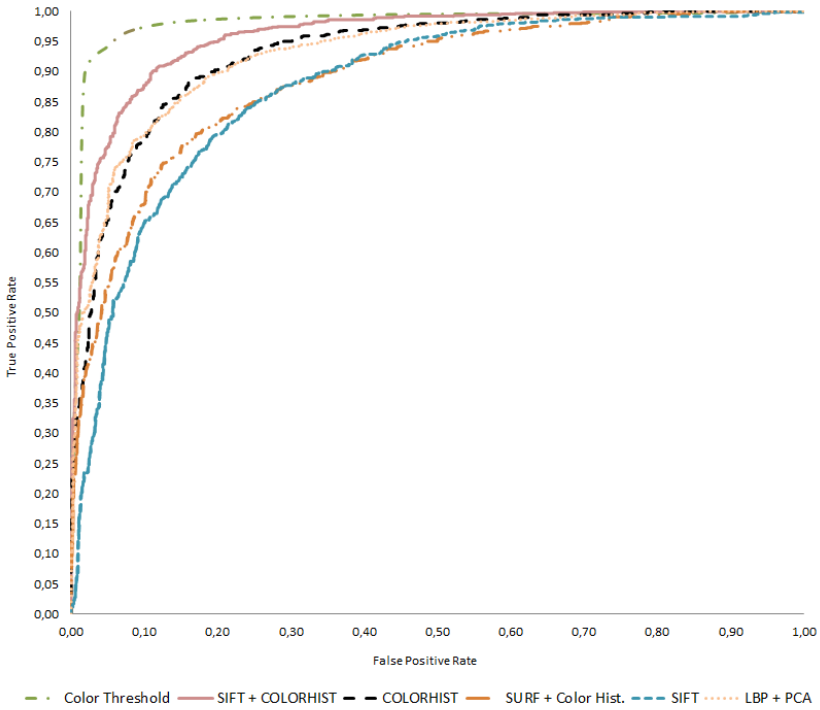


Figure 5.13. – ROC-Curve Showing Results of Classification with Normalization and AdaBoost Classifier.

regular texture properties like folds and shadow. In the feature extraction process of the LBP approach the selection of an adequate radius is furthermore error-prone, as the selection is very sensitive to the image quality. The local interest points SURF and SIFT are rotation and scale invariant but contain no color information. The combination with a extracted color-histogram tended to show better results in all approaches. The results indicate that local interest points features are not able to distinguish between shadow and stain defects (see Figure 5.3). The results of the approach using a simple color-range threshold via gray-scale conversion on the not-normalized textiles confirm that assumption. The classifiers SVM and AdaBoost achieved different error-rates using the same input data. For the calculation of the results using the proposed normalization as input data, only AdaBoost training was used. The scores using normalized textile images as input showed better results for all examined feature and classification methods. The approach using color-range thresholding resulted in the lowest error rate of 5.68%. A proportion of positives that are correctly identified of 95,28% (TPR) and a FPR of 4% were achieved using this approach. The ROC (Receiving Operator Characteristic) curve in Figure 5.13 shows the metrics TPR and FPR in relation to the threshold using the presented normalization technique and AdaBoost classifier. All approaches were evaluated using 4-fold cross validation. The Equal Error Rates (EER) per approach were defined by calculating the mean of all individual classifiers EERs.

5.4. Textile Defect Detection via Deep Learning

In this section, three methods based are presented and one, based on keypoints are presented. In the keypoints method, a localization of image regions important for recognition is proposed. This method is unsupervised and bases on SURF keypoints, that doesn't require any training data. I propose using their location, number and orientation in order to group them into geographically close clusters. Keypoint clusters also indicate the exact position of the defect at the same time. I furthermore compared our approach to supervised methods using deep learning presented in Sections 5.4.3-5.4.5. The presented processing pipeline shows how normalization and classification methods need to be combined, in order to reliably detect fiber defects such as cuts and holes. I evaluate the performance of our system in real-world settings with images of piles of textiles, taken in stereo vision.

5.4.1. Pre-Processing

For the later presented unsupervised methods to work, it is beneficial to normalize or free the images first from shaded or folded areas. The following section describes the pipeline of how to free the image of them in order to receive a more simple to interpret representation of it.

5.4.1.1. Textile Inspection Pipeline

Our complete inspection system pipeline (see Figure 5.14) classifies washed textiles in pile-like arrangement into the classes "defect" and "without defect". We suspect that the background, stain defects, shaded regions and folds have an negative influence on the detection of fiber defects. Therefore we experimented with pre-processing in which we generate masks of those regions (see Section 5.4.1.3).

The general architecture is built on a hierarchical decision tree model in which a first classifier (see Figure 5.14c) detects stains and prevents them from being further processed. In cases the amount of defected stains is below a certain threshold of acceptance, the second classifier (see Figure 5.14d) recognizes fiber defects. We present different methods that can be used in this second step of the pipeline, each of them shows different capabilities in terms of localization, flexibility to different fibers and training/annotation effort (see Sections 5.4.3, 5.4.4, 5.4.5 and 5.4.6).

The system is intended to be used in an assembly line like environment, where every item is served individually. The fabrics are captured from a top-down perspective, therefore, some defects might be hidden (e.g. if they are in a fold or on the bottom side). Iterations in which the textile is physically reoriented and then classified again could solve this problem.

5.4.1.2. Image Acquisition and Database

Our approach is evaluated on the textile pile image database version 2 presented in 5.1 (see also [SBK16] and [SSF*17]). It consists of dry cleaned woven cotton cleaning textiles as they are used in many different industrial applications. The best performing methods are also examined using the images in version 3 of the database.

5.4.1.3. Background and Shadow Masking

In this pre-processing step we segment the background from the foreground by using a mask. We convert the image into the HSV color space and define a color range for the 'H' (hue) value. The range that defines the

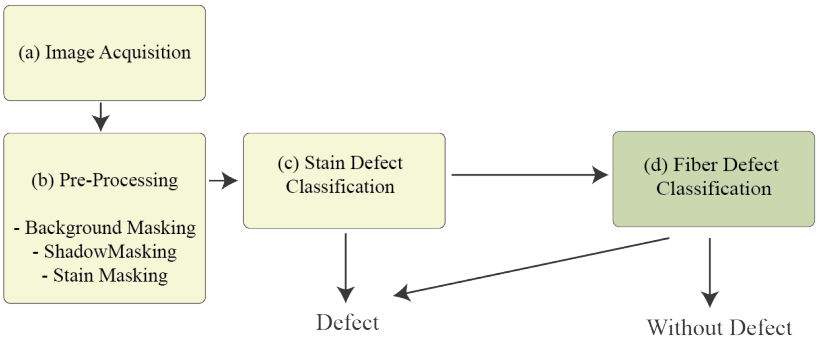


Figure 5.14. – Pipeline showing the order of the processing steps; Prep-Processing (b) and Classification Tree Steps (c) and (d).

background color can range from lowerH to upperH. Pixel in mask(I) are set to 255 if src(I) is within the specified range and 0 otherwise.

$$\texttt{mask}(I) = \texttt{lowerH}(I)_0 \leq \texttt{src}(I)_0 \leq \texttt{upperH}(I)_0 \tag{5.6}$$

We invert the mask and apply it onto the source image. The morphological operators erosion and dilation are used to exclude remaining smaller artifacts in the background from the foreground.

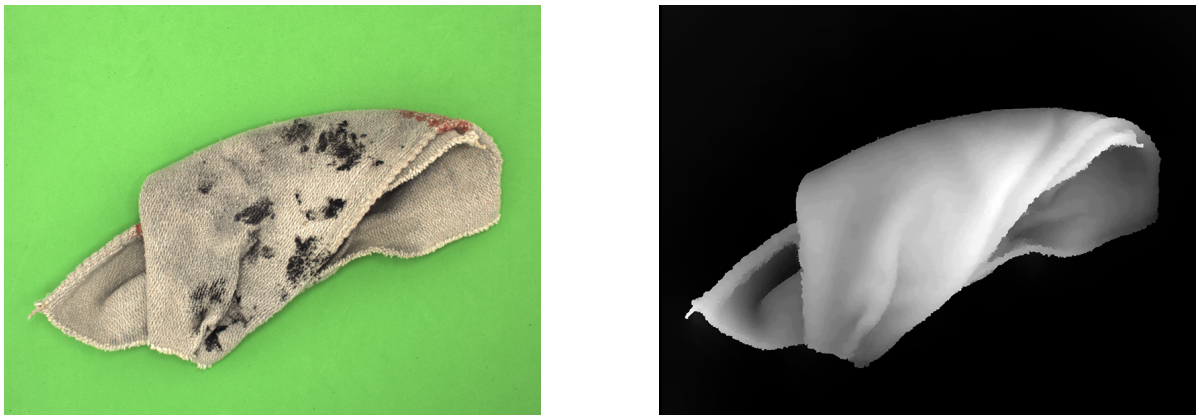


Figure 5.15. – (a) Input Image (b) Disparity Map.

In a previous work on the detection of soiled textiles [SBK16], we observed that shadow from ambient occlusion may influence the classification accuracy negatively. Therefore we experimented with eliminating such areas along folds in the same way. In that normalization process, we calculate and use the disparity depth map image for finding and normalizing of these areas (see Figure 5.15). First, k-mean clustering is applied to reduce the number of depth levels and recognizing folds. Second, the transitions along those folds are classified by using two features: the depth and color values in regions with folds. The depth feature is calculated by analyzing the absolute altitude between both sides of a fold using a sort of sliding window. The color value feature is needed in order to verify that only dark areas get detected. It is calculated by comparing both sides along a fold within

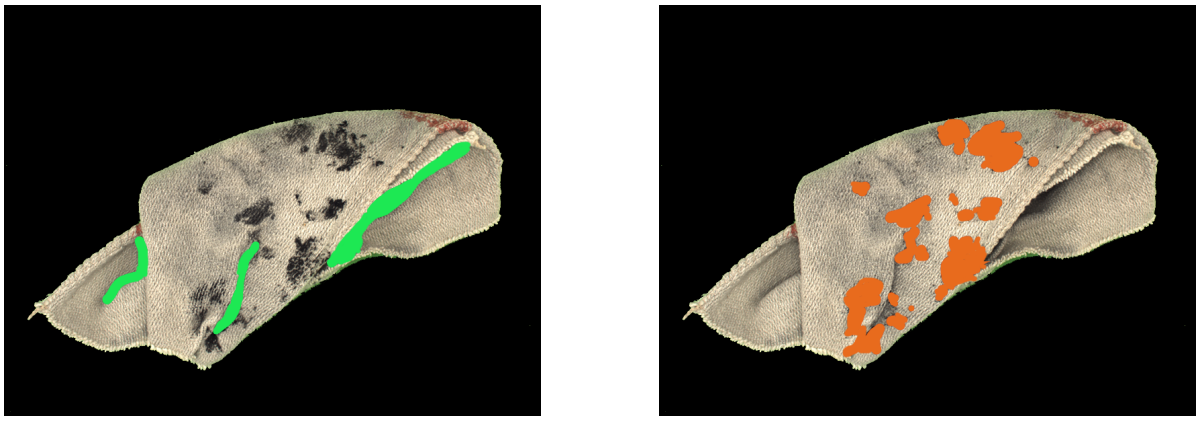


Figure 5.16. – (a) Shadow and (b) Stain mask (highlighted).

the range of the fiber mean color. For further processing, a binary mask {0,255} representing detected shadow is generated, that can be used to excluded these areas (see Figure 5.16a).

5.4.1.4. Stain Masking

Stain defects are characterized by the fact they are of different color compared to the fiber. Applying the shadow mask and the background mask to the original image, results in an almost homogeneous brightness of the image. Areas with a different color value from the fiber color can easily be detected with the aid of a grey-scale threshold range as shown in previous work [SBK16]. First, we convert the image into gray-scale and use k-mean clustering with k=2 on the masked image (assuming that defects are of the same color). The fiber mean color is the gray-scale cluster with the most pixel. Second, we apply the same thresholding method that we use in background masking by specifying a lower and upper threshold around that fiber mean color. Morphological opening and closing is used to create a binary mask {0,255} of the areas that aren't of the same color (see Figure 5.16b).

Stain Defect Classification The amount of non zero pixel in the binary stain mask is used as score for classification. We calculated the FPR (False Positive Rate) and FNR (False Negative Rate) by using all scores over the database as thresholds. We choose a threshold that classifies fewer textiles as defect then by using Equal Error Rate (EER). The reason for that is, that a higher FNR is acceptable if a FPR of 0% is achievable. In fact, these settings result in less textiles with stain getting classified as "defect" but in the same time, no images without defects get falsely rejected. We could see, that some of the false negative samples, show both fiber defects and stain and might therefore recognized as fiber defect during further processing. In our database textiles with stains and no other defects, were classified with a TPR of 95,33% by using the defined threshold.

5.4.2. Unsupervised SURF Keypoint Clustering

In this Section we introduce a novel method for classification of fiber-defects by defining clusters of SURF keypoints. It doesn't require any annotations or learning process and localizes the defected regions based on found keypoints. We experimented with applying and not-applying all masks generated in the previous section.

We apply SURF feature recognition with a minimum Hessian threshold of 500 to identify distinctive feature points distributed along the textile under examination. However we have to point out, that feature points are found

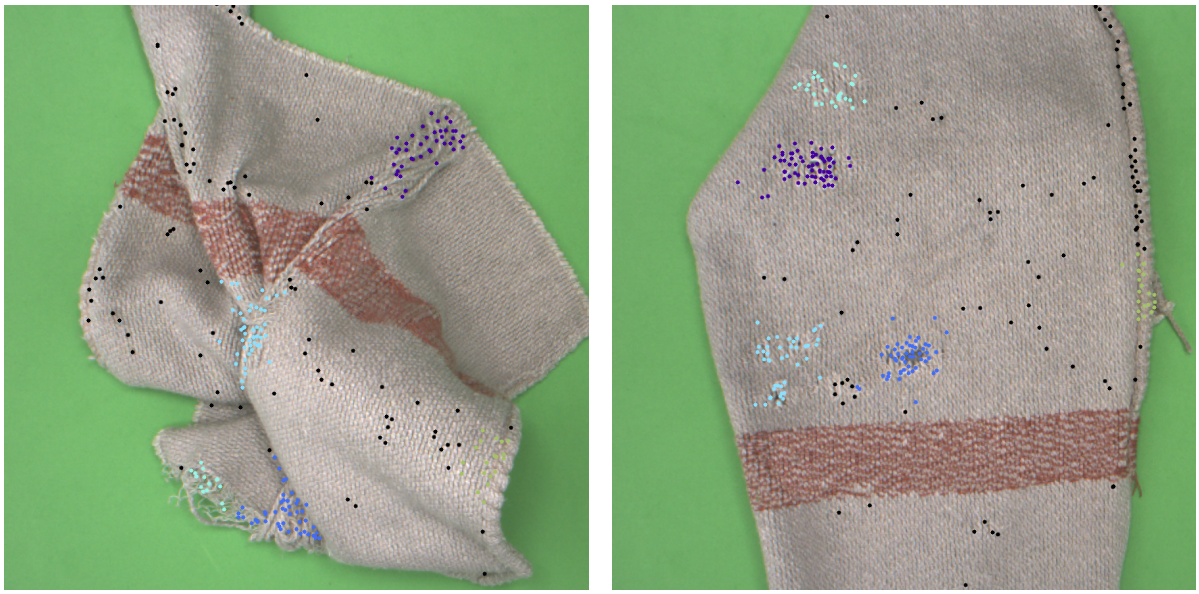


Figure 5.17. – Detected key points and DBScan clusters on textiles showing only fiber-defects. The black keypoints belongs to no cluster and will be considered as noise.

on the entire textile. Even after masking out the background, the previously identified stain locations and shadow regions (see Figure 5.17). Shadows are caused due to the random lying position of the textile on the conveyor belt. Using this presumably high threshold of SURF detector is to guarantee that only distinctive feature points will be found on certain dominant structures of the textile. We have tested different combinations of masks to detect keypoints which can be found in Table 5.11. The last detection step is to group the keypoints found into geographically close clusters using two clustering approaches: the partitional K-means [Mac67] algorithm and the density-based DBScan method [EKXS96].

The clustering procedure involves the partitioning of N data objects (keypoints) into a collection of K mutually disjoint subsets. Formally, let $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ be a set of N objects to be partitioned into K clusters $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_K\}$, such that the following three conditions are satisfied [TK09]:

- $\mathbf{c}_i \neq \emptyset$, $i = 1, \ldots, K$;
- $\mathbf{c}_1 \cup \mathbf{c}_2 \cup \ldots \cup \mathbf{c}_K = \mathbf{X}$; and
- $\mathbf{c}_i \cap \mathbf{c}_j = \emptyset$, $i, j = 1, \ldots, K$ and $i \neq j$.

Additionally, when partitioning the keypoints, the number of clusters is unknown *a priori*; therefore, the clustering problem consists in discovering the number of clusters K as well as the clustering $\mathbf{C}$ that best fits the actual data structure. To overcome this problem in K-means, the algorithm is run $m = K_{\max} - K_{\min} + 1$ times over the input keypoints varying the parameter K in the range $\{K_{\min}, K_{\max}\}$ in such a way that a collection of m clusterings is generated, where $K_{\min} = 2$ and $K_{\max} = 10$. Then, to select the final clustering solution, a cluster validity index (CVI) is computed for all the partitions in the collection and the one that obtains the best CVI value (maximum or minimum). The CVI is a function to quantitatively evaluate a clustering solution by considering the intracluster dispersion and the intercluster separation. In this work, the indices Silhouette (Sil) and Calinski–Harabasz (CH) have been used as the selection criterion in K-means [JGGF16].

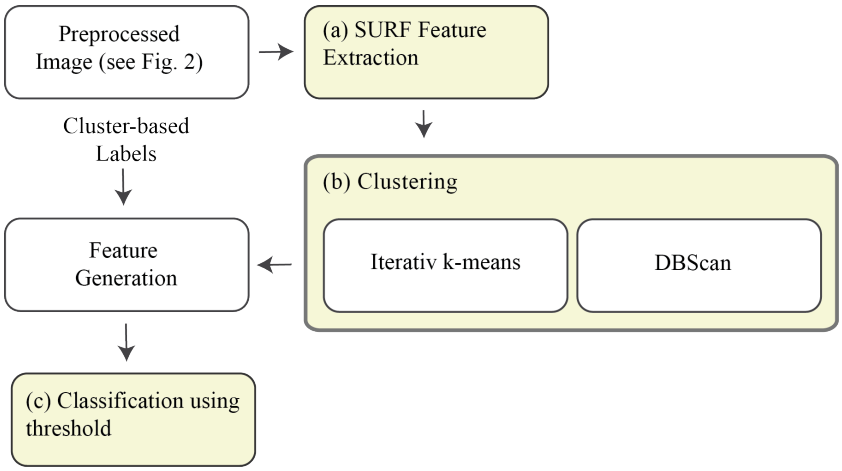


Figure 5.18. – Pipeline of cluster-cased defect detection.

On the other hand, the DBScan algorithm discovers automatically the number of clusters, which has proved to be useful when detecting clusters of different shapes and sizes. DBScan requires as input the minimum number of objects in a cluster (MinPts) within a radius (Eps). In our experiments, MinPts was fixed to 15 keypoints, whereas the parameter Eps was tuned for each dataset as follows: $Eps = \{0.1, 0.2, 0.3, 0.4, 0.5\}$, where each dataset was normalized in the range $[-1, 1]$ using softmax normalization [PK05].

After detecting the optimum cluster seperation, we now extract prominent features to distinguish whether a cluster indicates a fiber defect class or a non-fiber defect class. Our evaluations showed, that the cluster with included fiber defects contains distinctive attributes, which are the density and the feature point distribution within the cluster. This can be seen from the example image in Figure 5.17. These feature types are revealing the density distribution of the feature points and are used as indicator for fiber-defects. The feature we applied is the ratio between number of found SURF feature points N within a cluster towards the area of feature points distribution A using the Equation 5.7

$$r = \frac{N}{A} = \frac{N}{\sigma_x^2 \cdot \sigma_y^2} \tag{5.7}$$

within a cluster. The standard deviation $\sigma_{x,y}$ can be defined as in Equation 5.8

$$\sigma_x^2 = \sqrt{\frac{1}{N-1} \cdot \sum_{i=1}^{N} (x_i - \bar{x})^2} \tag{5.8}$$

and 5.9

$$\sigma_y^2 = \sqrt{\frac{1}{N-1} \cdot \sum_{i=1}^{N} (y_i - \bar{y})^2} \tag{5.9}$$

note that the position $(\bar{x}, \bar{y})$ indicates the center position of certain cluster C_i and N are the total number of found SURF feature points within this cluster. We found out that in case of a fiber-defect-class the number of SURF feature points are large and the distributed area A is small. They all are concentrated close to the fiber-defect and hence the ratio r is quite large in comparison to good textile parts. While on the other hand the number of found

SURF feature points is small and widely distributed among the entire textile and thus resulting in a large area A and small ration r. As regarding to the final classification results, this effect is clearly reflected.

Due to the linear separability of the feature, we use a simple thresholding to evaluate our data sets. We defined the ground-truth labeling using predefined masks of the defect region similar to our first methods (see Figure 5.18). If more than 5 SURF feature points lie within the defect region, this cluster will be declared as a fiber-defect-class, otherwise as a non-fiber-defect class. The test is conducted on all textiles.

5.4.3. SURF Keypoint Preselection and CNN Classification

This methods is using a combination of SURF key points and convolution neural network classification. We experimented with applying and not-applying all masks generated in the Section 5.4.1.3. The introduced method was previously published as a conference paper [SSF*17].

In conventional supervised machine learning methods, the labeling of data is a costly process. In order to reduce the effort to be spent, we use a SURF detector with a minimum Hessian threshold of 500 to determine key points on distinctive areas of the textile. We generate partially overlapping patches of 32 x 32 pixels in size, centered at each key point. These patches are then used to train a slightly modified LeNet-5 CNN [LBBH98] classifier. As a consequence patches of low dimensional data are generated, to be used in training of the CNN to recognize patterns. Instead of requiring an enormous amount of high dimensional data, we direct the network to key points of distinct areas of the image using SURF key points.

We labeled our database manually by defining a mask on regions with fiber-defects (see Figure 5.19). If a feature is inside the masked defect-region, its corresponding patches are classified as a defect patch. Patches outside that masked region are classified as a non defect patch.

We use two convolution layers which are combined with a max-pooling layer (see Figure 5.20 for a visualization of the customized network). After the inner-product layer we receive a fully connected layer to score our data. The loss layer is built using the soft-max function. To implement the network we use the caffe libraries from BVLC [JSD*14].

The so created database consists of approx. 58000 image patches and is divided in a training (80 %) and testing set (20 %). The data is separated and shuffled taking into account the individual images in such a way that no image is simultaneously present in the test and training set. It is trained with 1500 iterations and a batch size of 1000 features, which results in approx. 30 epochs. The classifier predicts whether a patch belongs to a region showing some fiber-defect or not. For each textile we receive as many predictions as there are key points.

In order to make a final decision for each textile, we use the weighted sum combination rule to calculate a fused unified decision. It is based on two features, the first one is the number of key points detected. The second feature the difference between positive and negative decisions. We represent both values as scores and normalized them to a comparable range using min-max normalization, which can be formulated as:

$$S' = \frac{S - min\{S_k\}}{max\{S_k\} - min\{S_k\}} \tag{5.10}$$

Where $min\{S_k\}$ and $max\{S_k\}$ are the minimum and maximum values of existing scores in the data of the corresponding sources and S' is the normalized score. We used the weighted sum score fusion, where for each score source a weight is defined that indicates its relevance on the fused decision. The weight is calculated by 1-EER and fused by the weighted sum rule F for N score sources.

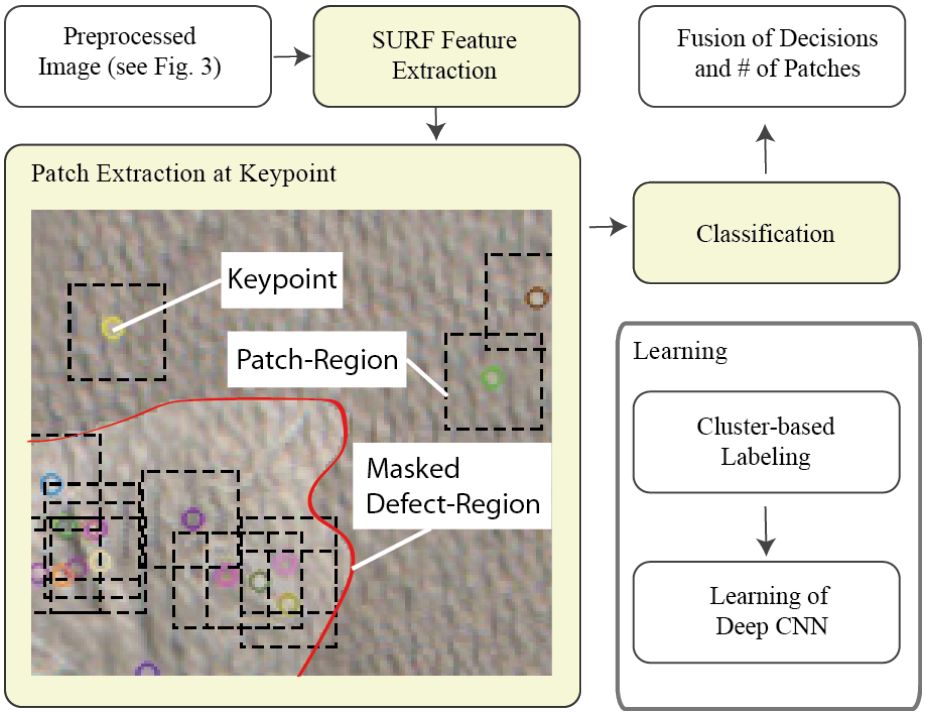


Figure 5.19. – Pipeline of Micro-Patch-based Classification using CNN.

$$F = \sum_{k=1}^{N} w_k S_k, k = \{1, \ldots, N\} \tag{5.11}$$

5.4.4. Deep Learning using Inception Modules

Following novel method was architecturally inspired of the proposed Google-LeNet [SLJ*15]. This approach is using a sliding window, where every patch of size 128x128 pixel was labeled manually and will be classified separately. The localization of a defect in the image is therefore restricted to that resolution. The architecture itself is quite different from a traditional CNN design like LeNet-5 model. In a typical CNN, convolutional layers are stacked together, each performing one convolutional operation at a time (e.g. 3x3 or 5x5 pixel). We used multiple inception modules, which perform multiple convolution operations and max pooling in parallel. Therefore its not obligatory to choose a certain convolution kernel size for a certain layer. This approach showed good results in ImageNet Large-Scale Visual Recognition Challenge 2014 and is not just efficient in classification results but also in computational efficient. Reason for the computational gain is 1x1 convolution operation which is applied before every 3x3 or 5x5 convolution of the inception module, it results in dimensionality reduction.

The purposed CNN design is adapted from GoogLeNet and modified towards compatibility with available hardware. In Table 5.9 the detailed architecture of the CNN design is described. Proposed architecture takes an rgb-image of 128x128 pixel as input. To avoid internal covariate shift, batch normalization is used for all convolutional and fully connected [IS15]. All weights are initialized with a normal distribution, using a standard deviation of 0.1 and zero mean. A batch size of 300 is used with the equal representation of both classes.

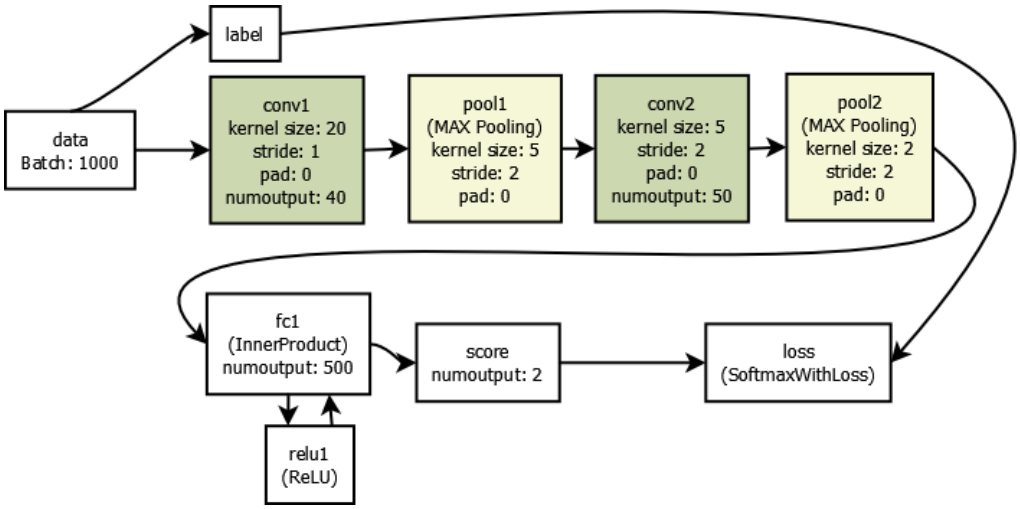


Figure 5.20. – Our adjusted LeNet-5 CNN.

Table 5.9. – CNN architecture based on GoogLeNet model.

Layer/Module	Input	Output	Filter, Stride
Conv 1	128x128x3	64x64x64	7x7x64,2
Max pool 1	64x64x64	32x32x64	3x3,2
Conv 2	32x32x64	32x32x64	1x1x64,1
Conv 3	32x32x64	32x32x192	3x3x64,1
Max pool 2	32x32x192	16x16x192	3x3,1
Incpetion 1	16x16x192	16x16x256	
Incpetion 2	16x16x256	16x16x320	
Incpetion 3	16x16x320	8x8x640	
Incpetion 4	8x8x640	8x8x1024	
Incpetion 5	8x8x1024	4x4x736	
Avg pool 1	4x4x736	1x1x736	4x4,1
Fully Connected 1	1x1x736	1x1x64	
Fully Connected 2	1x1x64	1x1x32	
Fully Connected 3	1x1x32	1x1x2	

The Training process ran for 200 epochs for each fold of the data set and each epoch contained 3 batches. As both classes are mutually exclusive, a softmax classifier is used with cross-entropy loss function. For all fully-connected layers a dropout of 0.5 is implemented. For stochastic optimization, an Adam optimizer [KA15] is used with a learning rate of 0.01. Except for the final layer, all layers including those inside the inception modules use ReLU activation. Sigmoid activation is used for the final layer.

5.4.5. Transfer Learning using VGG16 and Resnet

Following approach is following the recent publication [SPKK18] which is based on transfer learning and uses the Faster-RCNN architecture [RHGS15] in order to localize and classify regions inside an image as "fiber defect" or "non defect" (see Figure 5.21). It is fully supervised and needs defect image regions to be annotated. It reuses

CNN features for proposing regions instead of running localization separately through e.g. selective search [Gir15] or keypoints. Thus, the architecture consists of two networks, a Regional Proposal Network (RPN), that generates the region proposal and Fast RCNN that uses these proposals to detect objects (see Figure 5.11).

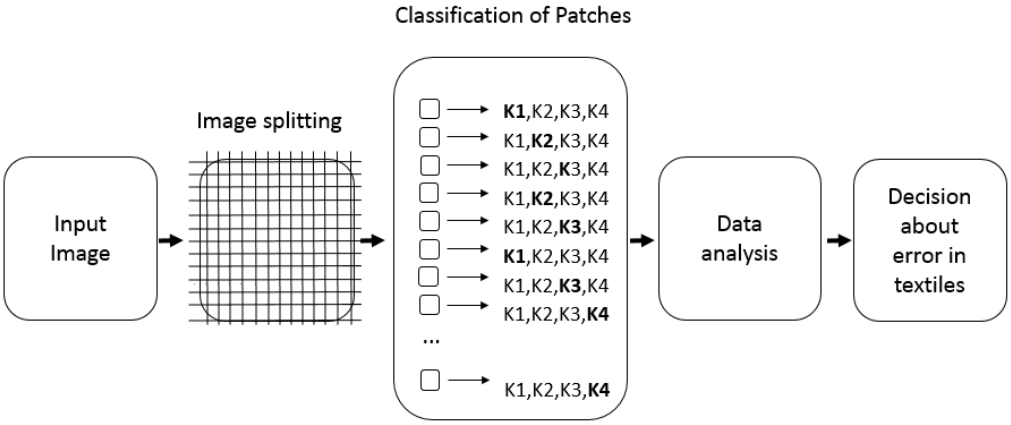


Figure 5.21. – Processing pipeline of the integrated neural networks.

The RPN is a fully convolutional network [LSD15], on top of the convolutional feature map. It works by sliding a small network (sliding window - n x n with n=3) over the convolutional feature map output by the last shared convolutional layer. RPN predicts numerous bounding boxes (k) at each locations. These k reference boxes are actually the anchors which is centered at the sliding window generating k = 9 anchors at each sliding position because by default three scales and aspect ratios are used. It can generate 300 proposals in less than 0.3s. Besides, Faster R-CNN also uses Region of Interest (RoI) Pooling layer that pools the feature map of each proposal into a fixed size (7 x 7). Thus it can handle defects of arbitrary sizes.

5.4.5.1. Feature Extraction Architecture

The residual networks "Resnet" [HZRS16] and the "VGG" [SZ14] network were evaluated to be used to extract features in the fabric in order to provide them to the RPN and Fast RCNN. Resnet has been evaluated on imagenet [KSH12] with a depth up to 152 layers which is 8 times deeper than VGG nets while still having lower complexity. VGG uses very small (3x3) convolution filters pushing the ConvNet depth to 16-19 weight layers. VGG uses a fixed-size 224 x 224 RGB image as it's input during training. It subtracts the mean RGB value in it's input image which is computed over the training images. The convolution stride is set to one pixel to preserve the spatial resolution of the convolutional layer. Similarly, it performs max-poling over 2 x 2 pixel window with stride 2 and as shown in the figure of VGG architecture, there are five max-pooling layer. Three fully connected layers follow a stack of convolutional layers, out of them two have 4096 nodes and the third with 1000 nodes performs softmax classification over 1000 class ImageNet database. All hidden layers use RELU as an activation function.

During testing we received consistently better results with VGG16 than with Resnet. That is why we used a pre-trained VGG16 (16 layers with 13 conv. layers and three fully-connected layer). We only use the first 10 convolutional layers (see Figure 5.22) and extract the features from the last layer which is passed to the RPN. The first four layers are freezed and the remaining are fine-tuned. Our design also uses the first 10 convolutional layers and two fully-connected layer, each with 1024 neurons and ReLU activation functions. The learning rate of 0.0001 and and then decay rate at 0.2 is used after 10 epochs.

As training of large input is computationally costly, we reduced the image size to a maximum of 600 pixel at the largest side. All presented design experiments use a dropout of 0.5 in the fully-connected layers and are trained using the AdamOptimizer. The training data was flipped horizontally, vertically and further augmented by adding artificial blur, contrast adjustment and sharpening. This resulted in 10,843 images available for training.

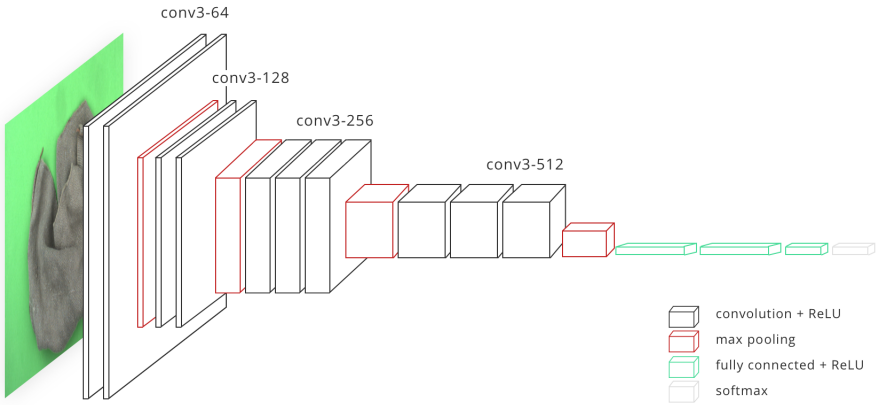


Figure 5.22. – Adapted neural network design, based on the VGG16 model.

5.4.5.2. Dataset

We evaluated this approach on an extended data-set of washed woven cotton cleaning textiles collected in the same way as described in Section 5.4.1.1. It consists of 2415 images (out of 306 textiles) with ground truth in Pascal VOC [EEVG*15] format. Figure 5.23 shows some samples of the defect class. The data-set is split into a train set of 1551 images and a test set of 864 images.

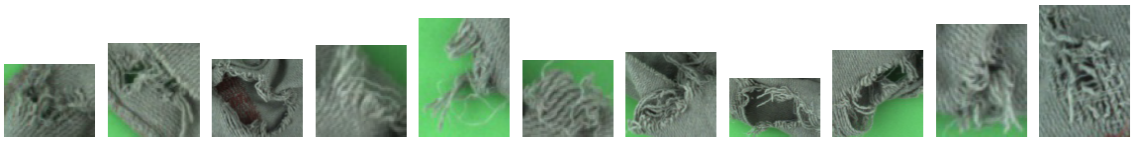


Figure 5.23. – Samples of fiber defects regions.

In order to remove noise from the background of the disparity map image, simple color masking was applied, as described in [SBK16].

5.4.6. Conventional Feature Classification

For comparison we applied the following conventional feature extraction and learning methodology. We used the same non-overlapping patches of 128x128 Pixel as in the previous Sections 5.4.4 and 5.4.5 in order to reduce the complexity of the analyzed pattern. Two different sets of features were selected from related work and examined as fiber defect classification step (see Figure 5.24).

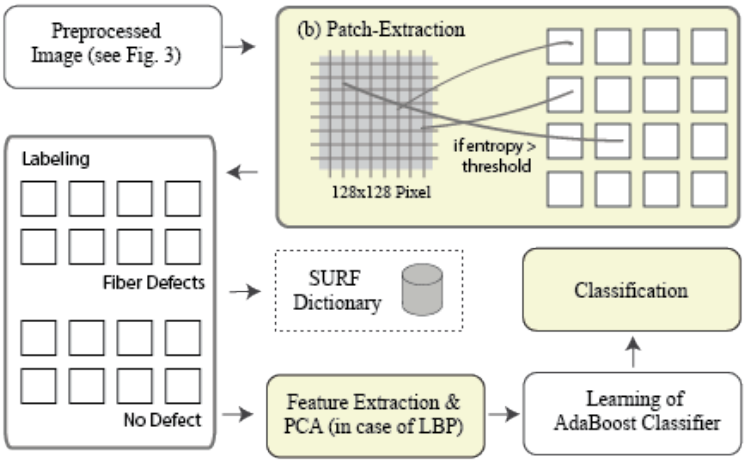


Figure 5.24. – Pipeline of Conventional Classification using Sliding Window.

During extraction, we used the Shannon-Entropy-Value to determine the amount of information in a patch. Our evaluation showed that patches with an entropy-value below 2000 do not contain enough information to be classified and are therefore rejected. Every patch is labeled manually by assigning them the class fiber-defect (see Figure 5.8a) or none-defect (see Figure 5.8b).

The local interest point descriptor SURF [BTVG06] has shown its effectiveness in many applications as local feature detectors and descriptors for non-rigid 3D objects [ZWWY16]. They are scale-invariant and robust against rotation, translation and changing light conditions. A set of interest points is extracted into a 64-dimensional feature-vector, following a Bag of Words (BOW) approach[SKH16]. Local binary patterns (LBP) features are a known technique, when dealing with textile fiber classification tasks [SBK16]. The used LBP type [ZP07] is invariant against rotation and pixel intensity variations and shows a relationship between a pixel and its neighborhood. It fulfills the requirements in regard to computational cost compared to other scale-invariant LBP approaches[LLYY12]. An evaluation on a subset of the database showed that a radius of 3 and a block size of 32 pixels is optimal for the given pictures. Histograms of rotation-invariant binary patterns are calculated and concatenated to a feature vector. To reduce the dimensionality of the feature vector, Principal Component Analysis (PCA), trained on a subset of the data set, is performed. Experiments showed that a reduction to 300 components gives best results. An AdaBoost based machine-learning classifier is used for classification of the extracted feature vectors, using the REAL boosting method with confidence-rated predictions. Our evaluations showed that this classifier performed best among different evaluated classifiers (SVM, Random Forrest, AdaBoost and JRIP).

5.4.7. Results and Discussion

We evaluated our proposed methods described in the previous Sections on our database (see Section 5.4.1.1) with different preconditions. The accuracy in Table 5.11 describes the recognition of fiber-defects on textiles of the following defect categories: Fiber, Fiber and Stain and None (see Table 5.4). Patches showing other defects were excluded from the database using the first (stain-defect) classifier in the inspection process. The achieved accuracy (TP+TN/$\sum$Total population) was calculated on the full-image-level, which concludes to the final decision on whether a textile contains a fiber defect or not. In the calculation of that accuracy value the results of the first (stain-) detection classifier is not considered (see Section 5.4.1.1.4).

We performed experiments in order to find out if and which pre-processing steps are necessary. In the first experiment (see Table 5.11 (a)) no other mask than the background mask was applied to the image of a textile. In the second experiment (see Table 5.11, (b)), masks representing the background and stain defects were applied to the image. In the last experiment (see Table 5.11 (c)): background, stain defects and shadows were excluded from the image. All approaches were evaluated using 4-fold cross validation with a regular distribution of defect-classes in each fold.

Table 5.10. – Achieved Accuracy Rates of Fiber-Defect Detection Methods on TPD V2.

Methodology	Patchsize	Technique	Classifier	(a) w/o Masking	(b) with Stain Mask	(c) with Stain Mask Shadow Mask.
SURF BoW	128x128	-	AdaBoost	**89.4%**	87.6%	88.0%
RI unified LBP	128x128	-	AdaBoost	83.6%	82.6%	**84.9%**
SURF Clustering	Full Image	K-means Sil	Threshold	**94.7%**	87.9%	86.4%
SURF Clustering	Full Image	K-means CH	Threshold	**95.3%**	84.7%	84.0%
SURF Clustering	Full Image	DBSCAN	Threshold	**97.3%**	90.6%	87.0%
SURF+CNN	F.I.+32x32	-	CNN	**90.9%**	89.0%	90.2%
Deep Learning	128x128	-	CNN	97.0%	96.5%	**97.8%**
Transfer Learning	Full Image	-	CNN	**95.5%**	94.3%	94.1%

Among the conventional feature methods, the SURF BOW approach in combination with the AdaBoost classifier showed the best accuracy of 89.4%. Though using TanTriggs illumination normalization appeared to us to be promising, the results seems to be less accurate than without using it over all tested methods. Rotation invariant unified Local Binary Patterns showed overall less accurate results than the bag of words approach. We suspect the reason for the bad results to be that the rotational invariant method that is less information conserving than the SURF BOW approach. Applying or not applying the different masks to the image didn't influence the accuracy very much. In SURF keypoint clustering, without the masks, more keypoints can be found. Especially inside color bar and folds regions, keypoints that belong to one cluster get deleted by the mask and thus do not allow the image of a coherent cluster. Keypoints are linked to x, y coordinates, therefore accurate localization and determination of the size of the defected region is possible. A disadvantage is the hessian value which has a high influence on the number of found keypoints. Therefore, when applying this method, the right hessian value needs to be specified on some sample images. In the iterative clustering method, the quality of the result can be determined by the validation index; the disadvantage is that the index has to be determined for each image in several interactions, which is computationally intensive. DBScan is less computationally intensive as it

automatically determines the number of clusters by their density. The accuracy of this unsupervised method is comparable with the results of our later method using a supervised learning methodology.

The approach using a combination of key point pre-selection and convolutional neural networks (CNN) classification achieved slightly weaker results than the SURF keypoint clutering method. The best total accuracy was achieved by that method gaining an accuracy of 90.9% when no other mask than background removal was applied. Our method using deep net with inception modules, trained on image patches of 128x128 px, showed the most constant results in all experiments with an accuracy around 97%. The best results were achieved when training using patches with all masks applied. The fact that the performance varied very little, leads to the assumption that the build model covers all characteristics of the fiber defects well, independently from using further pre-procressing. A disadvantage of this approach is its need for training and the low localization performance, as relatively big patches were getting used.

All images in version 2 of the Textile Pile Database (TPD) suffer of the probability that certain areas are covered because they lay on something. Therefore, we also evaluated the TPD 3 using the best performing classifier using four images of the same textile, recorded in free fall from all sides.

Table 5.11. – Achieved Accuracy Rates of Fiber-Defect Detection Methods on TPD V3.

Methodology	Patchsize	Technique	Classifier	(a) w/o Masking	(b) with Stain Mask	(c) with Stain Mask Shadow Mask.
Deep Learning	128x128	-	CNN	99.1%	98.5%	**99.5%**
Transfer Learning	Full Image	-	CNN	**99.8%**	95.3%	99.2%

We classified each image of each set of four images individualy, bue we considered a set of four images as one applying fusing logic on the classification result of each set. Using the best performing methods we could achieve almost perfect results with more then 99% accuracy using the images recorded in free fall.

5.5. Summary

A novel database and several methods have been presented in this chapter. We presented the creation of a novel database called the "Textile Pile Database" (see Section 5.1), which contains images of textiles captured in a pile-like shape. The textiles were in clean and dry condition, but still showing dirt, holes, or other defects. There are two versions of the database where Version 1 contains 436 images and Version 2 contains 910 images of 258 different textiles with and without defects, captured using two different cameras. The textiles were annotated manually based on their different characteristics, with the main difference between version 1 and 2 of the database being the inclusion of disparity maps in the latter. The textiles contain different characteristics such as holes, stains, and texture properties like shadows, edges, and folds, and were captured using controlled lighting and a black box to protect them from external light. The database will be used to classify textiles based on their different characteristics. Section 5.2 discusses the use of automated inspection for textile manufacturing in order to improve quality and reduce labor costs. It focuses on 2D methods for fabric classification, using the visual descriptors LBP and SURF in combination with the classifiers SVM and Adaboost. The results show that using full images resulted in better performance compared to using image patches and the SURF interest point features

performed better than the LBP feature. The evaluation was performed in Version 1 of the Textile Pile Database. The Multiclass SVM classifier also outperformed the Adaboost classifier by an average of 3.67% accuracy.

Section 5.3 presents a stereoscopic normalization approach that is evaluated in a recognition pipeline of stain defects and integrated with several object recognition techniques. In conclusion, the paper presents a study on the use of different feature extraction methods and classifiers for the automated detection of stains on textiles in pile-like arrangements. The study found that the combination of local interest point features with a color histogram tended to show better results than using LBP features with SVM and AdaBoost, which achieved poor performance. The results also indicate that local interest point features are not able to distinguish between shadow and stain defects. The proposed normalization technique using color-range thresholding resulted in the lowest error rate of 5.68% with a high true positive rate of 95.28%. The approach was evaluated using 4-fold cross-validation and the equal error rates were calculated by taking the mean of all individual classifier EERs. Section 5.4 presented in summary four novel methods with performance evaluation for recognizing fiber defects on textiles using the Textile Pile Database in Version 2. All methods achieved over 90% accuracy in detection of fiber defects. The best performing, completely supervised method achieved an accuracy rate of 97.9% using deep net with inception module methodology. Using deep neural networks with inception modules trained on image patches of 128x128 pixels showed the most consistent results across all experiments. This approach required training and had low localization performance as relatively big patches were used. The best unsupervised method achieved a comparable accuracy rate of 97.3% only in recognizing stain defects, but it is additionally capable of defining the exact pixel location and therefore, the size of a defected region. By using Version 3 of the TP database, the two best performing methods outperformed all other methods of laying textiles. The best results were achieved at 99.5%, using the transfer learning method.

6. Conclusions and Future Work

Throughout this thesis, our investigation has been driven by a set of five research questions outlined in Section 1. In Chapters 3, 4, and 5, we extensively addressed each of these questions, providing comprehensive insights and analyzes. Now, in this concluding chapter, we consolidate the key findings and implications derived from this research, while also offering a glimpse into potential avenues for future investigations.

6.1. Conclusion

In this section, we present the conclusive findings that address the research questions and gain valuable insights into the potential impact of these methods on practical implementations, offering a glimpse into a future where these applications can be realized with increased viability and efficiency.

6.1.1. Biometrics

One of the main research topics we raised was the question of which methods will facilitate the industrial use of biometrics. In Section 3.1 we discussed the application of biometrics in industry and focused on three technical innovations that aim to improve its accuracy. In particular, we identified the accuracy of the algorithms as one of the issues that reduce applicability. This has led us to take a deeper look into multi-biometrics, and more specifically MFP, to see if they could improve precision:

> **Research Question 1: Is melamine face pigmentation a biometric modality and can it improve conventional facial recognition algorithms?**

We examined three topics in the context of practical applicability to answer this question. The first two topics deal with face recognition and the use of melanin face pigmentation (MFP) to improve its performance. The third topic deals with the vulnerability of biometric systems against fake biometric characteristics, specifically presentation attacks.

> **Research Question 2: Are there novel feature types or modalities which can be used to increase PAD performance?**

We also examined the use of MFP as a presentation attack detection technique. The final topic addresses the problem of double enrollment in biometric systems by examining the recognition of similarities in handwriting using forms. We found out that in biometrics, melanin facial pigmentation is an underestimated property that can improve the accuracy of face recognition. We also demonstrated its usefulness for the use case of presentation attack detection. In the third example, we found that low-level features also contain valuable information that increases the applicability of anti-spoofing attacks that attempt to detect double enrollments.

Face Recognition We have introduced new biometric face features in the ultraviolet (UV) spectrum that improve the accuracy of current face recognition systems. These features can be added to existing systems by

capturing images of subjects in the UV wavelengths. By retaining the effective aspects of traditional face recognition algorithms and amplifying their accuracy with the extended information, the selection of useful data is also improved through the application of PAN-min-max and OLDW on the score level. This approach is considered widely applicable and has the potential to enhance the performance of current algorithms like OpenFace if they incorporate UV face images without altering their basic operation. With the advancement of neural networks and deep learning, the use of UV spectrum in face recognition has the potential to bring further performance improvements, presenting itself as a promising avenue for future research. We tested the data using local features, which we expected to find in high frequency features of sizes between 3 to 20 pixels. We created a dataset for the ultraviolet spectrum because a comparable dataset was not found. We evaluated the available front images from their database using three different algorithms (LBP, SURF, and CSLBP) and feature types. The algorithms were chosen for their differences in computational speed, rotation variance, and size tolerance. The best-case parameters for each experiment were used and the Euclidean distance between two vectors was calculated as the comparison score. The LBP features resulted in 359936 values, while the CSLBP features had 1600 features. Different feature types (LBP, CSLBP, and SURF) were tested on UV and VIS images using a "one versus all" approach. The results were measured using receiver-operating-characteristic curves and show an increase in performance when fusing scores. The LBP descriptor performs better on VIS images while the CSLBP descriptor performs better on UV images. The best-performing descriptors were selected and fused, resulting in a significant increase in performance in two of the analyzed approaches. The results also show that skin type has a high influence on the performance of the MFP feature, but due to the small size of the dataset, skin-type wise statistics of performance cannot be obtained. The results are shown using true acceptance rate (TAR) at a false acceptance rate (FAR) of 0.01 and 0.1. We suggest that the performance could be enhanced by considering feature level weighting and selecting descriptors based on facial features that are dominant under UV light, such as the UV-MP. Another approach could be to use localized descriptors, like CSLBP or LBP, on important facial locations, taking into account the nearest neighbors, to provide a more dynamic, interconnected, and structural description, as the static division into defined size matrices can introduce a lot of noisy information.

We incorporate that UV images in face recognition systems can improve verification performance. By retaining the well-performing aspects of traditional face recognition algorithms and enhancing the overall accuracy with extended information from the face, the performance of current algorithms could be improved. The application of PAN-minmax and OLDW to the score level enhances the selectivity of important data, which is considered generally applicable. Although the database cannot be released due to privacy reasons, we offer the opportunity to run tests.

Presentation Attack Detection We presented a study on the use of ultraviolet (UV) light spectra in face recognition, an increasingly common biometric modality. Despite the popularity of these algorithms, their security has been called into question due to their vulnerability to presentation attacks. This study is the first to explore the properties of skin using the UV spectrum for presentation attack detection. The authors used a multi-sensor approach that involved learning features from comparisons of visible and UV images, using brightness and keypoints as features and testing different learning strategies. The results of the evaluation were carried out on a novel face UV PAD database and showed an APCER/BPCER of 0%/0.2% in a leave-one-out comparison. The proposed method was tested on an extended version of a newly created UV-face database that includes images collected in both the UV and VIS spectrum in controlled conditions similar to border control. The database includes subjects of different ages, genders, and skin types and has been expanded to include 127 images of spoofing attacks using various materials. We created a new multispectral face image database of 91 subjects and selected presentation attacks based on reported successful attacks against face recognition systems. A combination of global and local features from ultraviolet (UV) and visible (VIS) images was used to determine if the image

was a bona fide or attack image. The results indicate that UV images in presentation attack detection contain useful information that is not easily overcome. The images were captured using two cameras positioned 1.5m away from the subject and illuminated by two 36W UV-A LPS lamps. The UV images were captured using a DLP LLC camera with a CMOS sensor, while the visible images were captured using a Nikon D9000 camera. All images were converted to grayscale and slightly changed in saturation before testing. UV images contain valuable information for presentation attack detection (PAD) and their method demonstrated a high level of accuracy compared to other PAD approaches. The results indicate that UV images contain useful information for presentation attack detection that is difficult to overcome.

Handwriting Identification In Section 3.3, we presented a method for detecting manipulations in anonymous offline questionnaires used for reviewing services or products. The approach involves a combination of features extracted from handwritten text, numbers, and checkboxes to identify duplicates created by the same reviewer. The nine extracted features undergo processes of alignment normalization, segmentation, feature extraction, classification, and fusion. The method combines different features of handwritten text, numbers, and checkboxes to detect such manipulations. A novel database containing pages of handwriting from 1,734 writers was used to test the proposed method. Nine features were extracted from the handwritten elements and used to recognize duplicates. We used various methods to extract features including writing zones, a color histogram, line width, checkboxes, digit height, slant, and SURF (speeded up robust features). Writing zones were extracted using connected components and hough transformation. The color histogram was calculated using the HSV color space. Line width was calculated using erosion and structural elements. Checkboxes were analyzed for their type of mark, length, position, and order of strokes. Digit height was examined using the date field. Slant was calculated by two different methods, - one based on the average angle of almost vertical elements and the other by calculating means of horizontal and vertical projection histograms. SURF was used to extract interest points from a set of images. The authors used a weighted sum combination rule to produce a unified biometric score. The system consists of five parts: alignment normalization, segmentation, feature extraction, classification, and fusion. The features are mainly based on visible characteristics of the writing, such as color, slant, word proportions, and crosses as well as SURF features. In the context of this research, a total of 1,734 questionnaires based on 34 different questionnaire types were used. To assess the results, four test subjects each filled in 10 questionnaires (4 times 10 Genuine, 1,694 Imposter). The EER (equal-error-rate) was calculated for each feature individually to demonstrate their significance. The results of the various feature methods and their classification method are shown in Table 2. The ROC (receiver operating characteristics) curve in Figure 4 displays the characteristic performance of the different features under different TPR (true-positive-rate) and FPR (false-positive-rate) prioritization. The color histogram and SURF features had the best performance with an EER of 8.2% and 10.7% respectively, compared to all other feature types (see Table 2). Local writing characteristics such as slant or writing zones are not as distinctive as color or interest points. Although these properties are weaker, the characteristics they represent are still important. The slant of writing from the "open text" field, when read using the approach by Bozinovic et al [BS89], resulted in a high error rate of 36.97%, while the vertical histograms of single words showed only 24.7% EER. This might be due to the slope of writing lines in the open text¨field, negatively affecting the distinctness. The analysis of digits and crosses is not based on letters and therefore adds complementary results to the multi-modal characteristics. By fusing all scores of the extracted features, an EER of 4.33% was achieved. The aim of this study was to create an effective method for checking handwriting duplicates in customer reviews. A unique and realistic database was gathered to facilitate the development and assessment of the approach. Several distinct features were analyzed and combined into a unified method. It has been shown that writer identification performance can be significantly improved with low-level features, thanks to the use of a weighted sum combination rule for making the biometric decision. The outcome of the solution was a duplicate identification rate of 95.67%.

6.1.2. Enabling Autonomous Entrance Control

Chapter 4 discusses a current topic of critical infrastructure protection that includes both digital and physical protection. Physical protection involves separating individuals when entering buildings through systems like turnstiles and mantrap portals, which only grant access if certain requirements are met. In the following chapters, we evaluated technologies that could be capable of enabling autonomous mantrap portals. We outlined several methods to detect humans in restricted areas, including thermal imaging, RGB-D images, optical flow, pervasive capacitance, and multi-sensor solutions. All approaches have been analyzed in order to determine their reliability in recognizing piggybacking and tailgating and their potential for industrial application. Currently, solutions used in industry are not foolproof and involve simple sensors that can easily be circumvented. There are few studies available to evaluate the security of these systems, and determining the most useful sensor technology is a challenge. The goal of this research was therefore to answer the question:

Research Question 3: Which technologies are suitable for autonomously detecting piggybacking and tailgating when accessing restricted areas?

The methods and technologies described in the Sections 4.2-4.5 do not allow a clear unique answer to the question posed, as all technologies have advantages and disadvantages. However, the approach described in Section 4.6 draws on multiple technologies, one image-based and one sensor-based. This method has been found to be the best method, with only minor variations in the image-based methods. The feature that makes the difference is the mix of different technologies and different imaging positions. One certain result, however, is that a single technology is not sufficient to ensure secure autonomous access. Secondly, the technology used should be multi-modal, i.e., a single sensor that measures only one modality is not sufficient. To arrive at this statement, we have used the following evaluation scheme where we used a selection of different people, sexes, and body measurements. Two different knowledge levels were also tested to simulate prior knowledge. Tests were also carried out in which the participants were supposed to take objects with them into the portal. We then presented four novel image-based methods and two capacitive sensor-based grids, each trained and used with machine learning. Each of the methods used has shown strengths and weaknesses. The thermal approach 4.2 suffers from a high false positive rate when people take aids into the portal or use them to manipulate the camera. Nevertheless, the evaluation of the presented system shows that thermal imaging is a useful technique in verifying the isolation of people, especially if pose and positioning guidelines (like standing upright, staying still) can be established. False color RGB-D images are another technique, presented in Section 4.3. The results show very low EER in the case of verification with an identity claim. The results show that there is no scenario in which an attacker can be sure to overcome the system. However, there are sensor limitations which reduce its practical usability. The method using optical flow 4.4 performed better than the other two methods. The worst detection rate was also achieved in breach attempts in which the attackers used aids to breach the portal in a targeted manner. The last presented imaging approach 4.6.5 uuses background subtraction on an image series and deep learning for classification. This approach achieved the best results of all approaches but requires the people to stand still at a marked position. A capacitive sensor presented in Section 4.5 tries to detect people sneaking in behind authorized individuals to pass through the transit space (tailgating attacks). Objects brought along, such as suitcases or cleaning equipment, which are occasionally misidentified, cause problems here. In Section 4.6 we showed that a combination of the sensor grid and the imaging method based on deep learning achieved the best results and is able to recognize all piggybacking attacks. A limitation of the system is the requirement of the user to place himself on a position marked on the ground, since the performance otherwise decreases strongly. In the next Sections, we conclude all four different approaches examined, including thermal imaging, RGB-D images, optical flow, and measuring the pervasive capacitance of feet. The result of each approach is analyzed to determine their effectiveness in recognizing piggybacking and tailgating, and their suitability for industrial use.

6.1.2.1. Thermal Imaging

In this method, we used thermal imaging to detect humans based on their body heat and its performance was tested by exposure to various scenarios involving changing appearances and objects carried by the individuals. The results of the tests were analyzed using receiver operating characteristic (ROC) curves. With respect to the TPR of the evaluated system, we can summarize that objects have a measured impact of +/- 10% on the TPR. Whether objects radiate heat or not does not affect the measurement accuracy. The size of a carried object only has an impact on the system if the test subject cannot position themselves adequately in the mantrap. With respect to the FPR, it can be summarized that the height or width of a person exerted no detectable influence. Individual attack scenarios have very different success factors. Attacks which, although carried out with prior knowledge of the system used, but without aids, have very little chance of success. Attacks carried out with aids achieve an average FPR of 30%, with a mirror by FPR affecting the system the most. The environmental parameters of light and temperature have different effects on the function of the separation system. While light does not affect the system, temperatures above 34°C result in significantly lower measured values that make the system unusable. The evaluation of the detected control images showed that user instructions were missing. In the tests, participants (test subjects) were able to choose their own posture and positioning inside the mantrap, which caused non-standard results.

6.1.2.2. Optical Flow

We introduced a single camera system for a mantrap which verifies that only one person is in the designated transit zone. Our unique solution combines optical flow and machine learning classification. A database of images of both normal verification and attempted attacks was created to test the system. Our novel approach for identifying attacks in an autonomous access control system is combining different approaches to human motion recognition. We have identified scenarios in which attackers tried to pass through a mantrap portal and classified them using optical flow features and machine learning techniques. Our method used dense optical flow which performs well in terms of motion recognition [GKWS09, DTS06] and enhanced its capabilities using micro movements, local descriptors, and variance calculations. Our assumption that this method would perform better in an attack scenario compared to the other methods and technologies using thermal and RGB-D imagery [SHK16, SWB16] proved the relevance of motion and micro movements in such a use case. We also noticed a positive impact using fusion. We provided competitive results and outperformed detection rates in various attack scenarios.

6.1.2.3. RGB-D

RGB-D images employ features that are invariant to rotation and pose, allowing real-time classification. We presented a pipeline in which images of an RGB-D sensor were used to analyze whether they show a single subject or a spoofing attack. A challenge in this task is the correct classification, even if subjects are carrying items with them (such as vacuum cleaners and cash boxes) or change their appearance in any way. With respect to speed and accuracy, low level histogram features were used and classified using distance metrics and machine learning techniques. We can summarize that the presented system shows some advantages to other approaches using weight or thermal imaging. The results show furthermore that there is no scenario in which an attacker can be sure to overcome the system. The performance of the system was evaluated by subjecting it to attack scenarios

and comparing it to existing methods. The results indicate that the presented approach outperforms competitors. The archived scores indicate that the approach is able to distinguish between humans with objects and

attacks in which an attacker is using additional aids, in most of the cases. The used sensor shows disadvantages regarding the minimal distance between sensor and subject (400mm) in the portal which is a limitation.

6.1.2.4. Grid of floor mounted capacitive sensors

Our new method takes inspiration from recent advancements in indoor localization using capacitance and measures the pervasive capacitance of feet in the transit area to detect tailgating. The presented method proposed using a grid of capacitive sensors. We investigate the ideal sensor techniques and grid layout for this application and used machine learning for classification. We explored suitable sensing techniques and its corresponding sensor hardware by combining well-performing aspects of known methods in the field of indoor localization. We identified attack scenarios, in which attackers tried to pass through our system and explored machine-learning classification strategies to identify them. Our method's performance was evaluated through hardware-level tests and comparisons to existing camera-based methods. With an equal error rate of 3.5%, our approach outperformed other methods and highlights the usefulness of combining it with other imaging techniques. Our evaluation proved the layout and performance of the proposed sensor grid, even under different environmental conditions. The performance of our method was defined in empirical testing, where we achieved good results in test-cases with feet of all subjects on the ground, even when only data of 0.5 seconds was classified. We assume that a combination of our method with an image-based approach focusing on movements like [SFS*16] will provide even higher security. Our method is vulnerable in cases where people are standing on one foot only, but it might be complicated for an attacker to lift one foot while standing still.

6.1.2.5. Multi-Modal approach using image and floor-mounted sensors

We presented a solution for preventing "piggybacking," using a combination of floor-mounted sensors and camera shots to verify the number of individuals within a designated transit area. Our evaluation proved the layout and performance of the proposed interactive sensor-grid and recognized all piggybacking attacks when combined with the image-based method. An image-based approach using change detection and a convolutional neural network was used to classify the motion extracted from a sequence of images. The solution is computationally efficient and capable of achieving state-of-the-art results individually. When merged, the solution effectively prevents piggybacking with a low rate of false positive results (BPCER of 7.1%). A limitation of the system is the requirement of the user to place himself on a position marked on the ground, since the performance otherwise strongly decreases. Both presented methods do not require high computing power and can therefore be used on single-board computers.

6.1.3. Industrial Inspection of Textiles

The reuse of textiles helps lower costs and is environmentally friendly, however, there are inefficiencies in manual inspection for identifying defects after washing and cleaning work clothes and industrial textiles. Automated visual inspection could provide a more reliable and stable solution for defect detection in this area. However, its implementation still suffers from challenges such as the similarity of global and local shadows and dark stains. The need to examine a wide range of fabric surfaces, the variations in fiber width and appearance due to the voluminous shape of textiles, and the fact that a defect can take any form or color make good generalization a hard job. Previous research has attempted to address these challenges through decision tree classification and preprocessing techniques, but the challenge of achieving high generalizability while maintaining a high recognition rate still persists.

We presented six methods for recognizing fiber defects on uniform textiles with a voluminous, pile-like shape, using four different types of woven cotton textiles as experimental objects. A novel database, named the "Textile Pile Database", has been introduced. The database contains images of textiles arranged in a pile-like shape, captured in clean and dry conditions but still showing defects such as dirt and holes. There are two versions of the database, with Version 1 containing 436 images and Version 2 having 910 images of 258 different textiles, captured using two different cameras and annotated manually based on their characteristics. The main difference between the two versions is the inclusion of disparity maps in Version 2. The textiles were captured with controlled lighting and a black box to protect from external light, and the database will be used to classify textiles based on their characteristics. These cloths were used throughout the study and defects were caused by heavy usage and the washing/drying process, including stains, bonding, silicon relics, holes, enclosures, dropped stitches, and press-offs. The use of automated inspection for textile manufacturing to improve quality and reduce labor costs was discussed in Section 5.2. The study focused on 2D methods for fabric classification using the visual descriptors LBP and SURF in combination with the classifiers SVM and Adaboost.

The results showed that using full images performed better than using image patches, and SURF interest point features performed better than LBP features. The evaluation was conducted on Version 1 of the Textile Pile Database, and the multi-class SVM classifier outperformed the Adaboost classifier by 3.67% accuracy on average. Section 5.3 presented a stereoscopic normalization approach for the recognition of stain defects, which was integrated with several object recognition techniques. The results showed that the combination of local interest point features with a color histogram tended to perform better than LBP features with SVM and Adaboost, which performed poorly. The proposed normalization technique using color-range thresholding resulted in the lowest error rate of 5.68% with a high true positive rate of 95.28%, evaluated using 4-fold cross-validation. In Section 5.4, , four novel methods for recognizing fiber defects on textiles were presented and evaluated using the Textile Pile Database Version 2. These included: (1) an unsupervised approach based on clustering SURF keypoints with an evolutionary algorithm, (2) a combination of SURF keypoints and neural network classification, and (3) a deep learning methodology trained using inception modules and patches extracted by a sliding window. All methods achieved over 90% accuracy in detecting fiber defects. The best performing supervised method achieved an accuracy of 97.9% using deep neural networks with inception modules trained on 128x128 pixel image patches. The best unsupervised method achieved an accuracy of 97.3% in recognizing stain defects, but it also had the ability to define the exact pixel location and size of the defected region. Using Version 3 of the TP database, the two best performing methods achieved a high accuracy of 99.5% using transfer learning.

In order to be able to answer research questions 3 and 4, we have had to carry out experiments on both the algorithm and the experimental setup.

Research Question 4: Is an algorithm able to generalize enough to detect defects on different materials?

Research Question 5: Which system setup in combination with which algorithms allows an evaluation in a comparable time compared to humans?

The analysis of the results of the different methods and especially the comparison carried out in Section 5.4.7 let us answer the raised research question 3 as follows. The deep net with inception modules showed the most constant results with an accuracy around 97%, and the best results were achieved when training with all masks applied. Masks did not have a significant impact on accuracy. The best results were achieved with the deep net with inception modules trained on 128x128 image patches. The method showed constant results with an accuracy of around 97%. However, to achieve the best results, it is also necessary to change the setup. Thus, the best results could not be achieved lying down, but in free fall. The database version TPD 3 has allowed

this evaluation using the best performing classifier, and almost perfect results were achieved with more than 99% accuracy using images recorded in free fall. The positive results of this system setup resulted in a software, called "'Classifly"', that has started to be used commercially (see Figure 6.1).

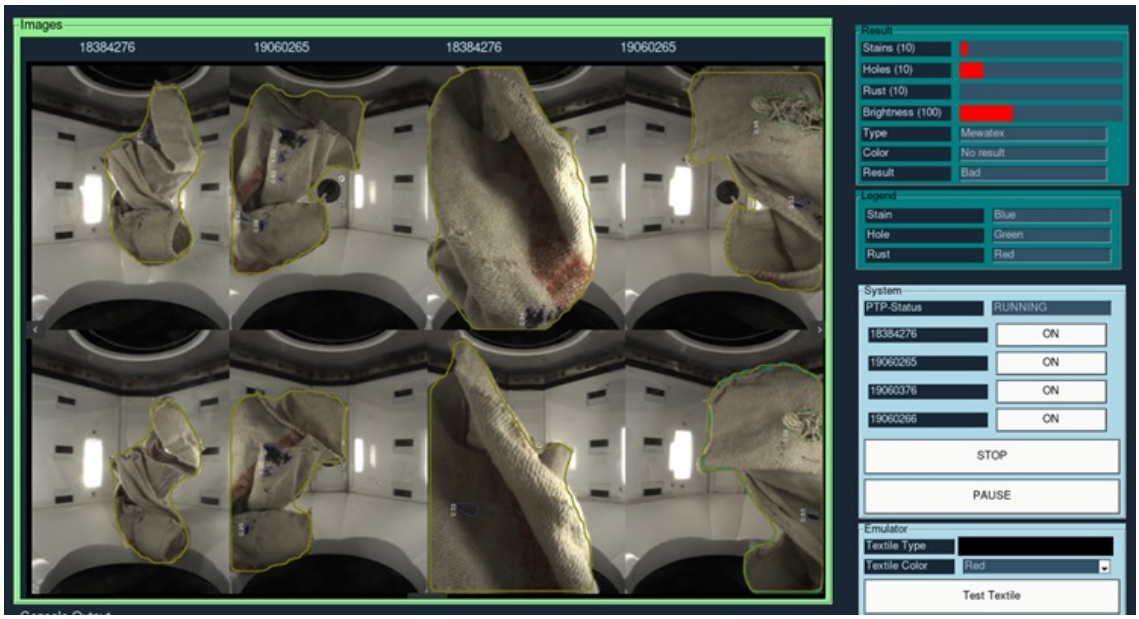


Figure 6.1. – (Interface of the Software "'Classifly"' showing Results of a Classification in Free-Fall.

6.1.3.1. Texture Interest Point Descriptors for Textile Defects

In this work, fabric patterns were classified using a database of textiles in a pile-like arrangement. There are multiple steps for classifying the fabrics: one involves extracting the features of woven fabric images; the other involves recognizing the class of woven fabrics. In order to find a solution which takes into account speed and accuracy, an approach which used patches instead of the full image was decided upon. Interest points as well as texture-analysis-based features were deployed and evaluated using different classifiers. For both identification and verification, the interest-point-based descriptor, SURF (in combination with multi-class SVM classifier), demonstrated the best performance. The patch-based approach reduced the calculation costs needed for prediction by 46% while showing reduced the accuracy of the verification by 3.67%. With the development of further methods, the image automatic identification and classification of woven fabrics could promote the development of the textile industry.

6.1.3.2. Stereo image Normalization of Textiles Arranged in Piles

We presented a novel approach for normalization of textiles arranged in piles, towards their classification regarding stain defects. Competitive approaches deployed on flat, and spread-out 2D textile surfaces were selected for an evaluation of this novel computer-vision application. The new database shows textiles in a pile-like arrangement, recorded with a stereo vision camera setup. For detection of stains in the textiles, machine learning as well as color-range thresholding classifiers were used. We showed that all evaluated features have disadvantages when

distinguishing between defect characteristics and regular textures of textiles arranged in piles. The presented normalization methods resulted in better classification results for all examined feature types. Nevertheless, the best results were achieved using a simple color-range threshold on the normalized fabrics. We described underlying assumption in Sections 5.4.2 and 5.3.5 and proved their relevance towards a robust classification of voluminous textiles.

6.1.3.3. Deep Learning Based Methods for Inspection

This study presented four methods for localizing and recognizing fiber defects on images of woven cleaning textiles after washing. All methods achieved over 90% accuracy in detection of fiber defects and can be used if textiles have a voluminous shape, i.e., if the images contain folds and occlusion. The best performing, completely supervised method achieved an accuracy of 97.9% using deep net with inception module methodology, where several pre-processed masks were used to reduce the dimensionality of the textile image. In this approach, patches of 128x128 pixels were used to recognize and localize defect areas. Our best model surpassed existing techniques based on stereo vision and even on images with conditions almost imperceptible to the human eye. Our unsupervised method achieved a comparable accuracy of 97.3% but is additionally capable of defining the exact pixel location and therefore, the size of a defected region. This method does not depend on labeled training data, such as other approaches based on neural nets. Our method using a combination of keypoints and CNN classification combined useful aspects of both approaches by using low dimensional data as an input. The presented processing pipelines show how normalization and classification methods need to be combined in order to detect different kinds of textile defects. We evaluated the performance of all methods in real-world settings with images of piles of textiles, taken using stereo vision. The effectiveness of our methods is described and proved in Sections 5.4.2-5.4.6 while a discussion and description of the achieved performance is presented in Section 4.5.3.2.

6.1.3.4. Summary

Figure 6.2 links the research questions to the corresponding chapters along with the publications.

6.2. Future Work

This thesis analyzed three computer vision applications and presented novel methods and databases that should make them more applicable. Nevertheless, these applications are very demanding, and thus prompt several future research directions based on the contributions of this thesis. These research directions can be summarized as follows:

6.2.1. Biometrics

The results obtained show that melanin face pigmentation is a useful property that can improve accuracy in biometrics as well as prevent attempts to overpower it. However, ultraviolet light has some fundamental disadvantages, such as a negligible proportion of this spectrum in artificial light. Together with its property of not being transmitted by glass, this makes it difficult to evaluate these properties indoors. The spectrum analyzed in the study was therefore in the upper wavelength range of UV-C radiation, which for skin and eyes is completely harmless when used for a short period of time (e.g., by a flash of lightning). If wavelengths in the even shorter range (UV-A or UV-B) were used, however, there could still be further or better skin properties. Here it would

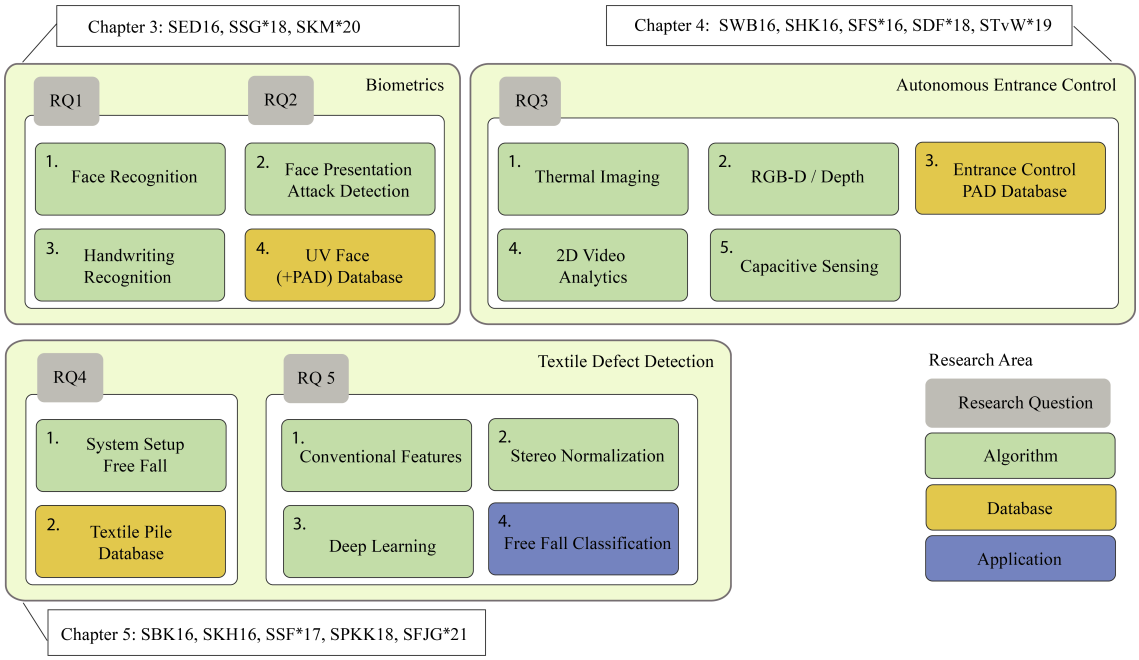


Figure 6.2. – A summary of contributions in relation with the research questions posed in this thesis, the chapters responding to those questions, and the publications building to these chapters.

be useful to have more sensitive sensors available, which would produce sharp images even at very low levels of this light in the scene to still not have to present a danger. By retaining the well-performing aspects of traditional face recognition algorithms and enhancing the overall accuracy with extended information from the face, the performance of current algorithms could be improved. We believe that future advancements in neural networks and deep learning could lead to even greater performance increases in the extraction and use of MFP features. The approach we presented has yielded good results, but the precision could certainly be further improved. Research to see if methods like Autoencoder or Transfomer could provide better results here might be helpful. In principle, the database would also have to investigate more test subjects with different skin types. A representative analysis would be desirable here. In this context, the generation of synthetic data could also be interesting, e.g., to enable the training of an even better detector.

6.2.2. Entrance Control

The results presented indicate a tendency as to which technologies are particularly suitable for certain use cases. Each of the technologies examined and each of the methods has individual advantages and disadvantages and can possibly be improved. However, if we look at the overall system and not at a single technology, further investigation with more test subjects would be advisable. The chosen approach of initially leaving the test subjects in the dark about how the monitoring system works during data collection and then specifically revealing information about how it works has already raised the creativity of the test subjects and thus the difficulty of the overcoming attempts to be detected very high. However, it would be interesting once again to test the solution ultimately described as the best, consisting of the sensor grid and computer vision once more with a larger group of people. Here a larger horizontal distribution of the persons would offer itself. Children and older people, as

well as more luggage, should also be examined, as is the case in practice. This would be suitable, e.g., in the context of a cooperation with a manufacturer from the industry. However, the individual technologies can also be further developed. In the case of the approaches in chapters 4.4-4.6, for example, it would be worthwhile to investigate them again with poses (e.g., the person must stand upright) and position rules (e.g., the person must stand in the center). The fact that the test subjects had no guidelines here made it difficult for the system to achieve very good results. A simple user default as it was applied later in the investigation could increase the performance again. However, this investigation will probably not lead to any revolutionary improvements in the results obtained. For the method with the RGB-D sensor from section 4.4, it might be worth trying other RGB-D sensors or stereo vision to avoid the limitations due to the minimum distance to the sensor. However, in order to develop the proposed solution of combining capacitive sensors in the floor and an algorithm based on a top-mounted camera to market maturity, the price of the capacitive sensors must come down and they must be available as a product. Alternatively, using resistive sensors would be a way to save costs. This technology has already proven itself in the area of multi-zone safety mats but requires that all feet are on the ground.

6.2.3. Textile Defect Recognition

Although the detection of defects in free fall worked well, based on the results obtained, this solution still requires that the defect is visible at all. As our study found, four cameras see only 61% of the cloth in free fall. It would therefore be interesting to know how letting a cloth fall through the inspection box several times could improve the quality of detection. Which settings and thresholds would be necessary to achieve the best possible performance in the end? Another interesting aspect of the system is that a balanced setting of thresholds is not necessarily the most sustainable solution. Thus, even with the best algorithm, false positives could be detected, i.e., stains on cloths that have no stains at all. With 1 billion cloths washed per year, false positives lead to the disposal of perfect goods. Therefore, a setting must be selected that prevents the detection of false positives, which, on the other hand, will also lead to a reduction in the detection performance of defects. A defect does not necessarily mean that a textile can no longer fulfill its function; there can be fine gradations. Therefore, the size of the defects plays an important role. It is important to know how big each pixel is in an image. In the case of the free fall box application, this calculation is very difficult, because the throughput of cloths is very high, leaving little time for calculations. Furthermore, space limitations do not allow the use of all theoretically possible sensor techniques. In particular, the distance between camera and object, a minimum of 5cm, is a major limitation. The use of stereo images is also not possible due to the spatial size. There are different ways to determine the size of textiles with the help of cameras. As a rule, however, this requires the depth to be determined for each pixel of the image. Established methods often use stereo images or the projection of a pattern to generate a point cloud in which the depth information is available for each point (rgb-d point cloud). Neural networks offer the possibility to compute information from image data with less mathematical effort, in a short time. In a future work, we would like to investigate if depth information can also be generated from a single image for this use case and if synthetic data can be of help in this case. Methods to determine the depth of objects on single images use so-called "visual cues", which can be extracted by convolutional neural networks and used to train a model (e.g., the work of Zhen et al. [ZCC18]). These are determined and used in the form of local, global, and super pixels as features of the ground truth. Existing algorithms could be examined for their usability with respect to the requirements mentioned. If necessary, further research to use synthetic data has to be found. The questions would be if there are methods that can be applied to generate depth information from single images and how precisely. To ensure real-time detection, it is important for all processing steps to be fast and simple, leading to further research questions in this field. In future work, we want to focus on this topic. Another future goal is to evaluate the presented methods on a bigger database and also with other types of textiles such as knitted fabrics.

A. Publications, Patents and Talks

The thesis is partially based on the following publications and talks:

A.1. Publications

1. SIEGMUND D., BRAUN A., KUIJPER A.: Stereo-image normalization of voluminous objects improves textile defect recognition. ISVC 2016, Las Vegas, NV (2016)
2. SIEGMUND D., EBERT T., DAMER N.: Combining Low-Level Features of Offline Questionnaires for Handwriting Identification. Springer International Publishing, Cham, 2016, pp. 46-54.
3. SIEGMUND D., FU B., SAMARTZIDIS T., WAINAKH A., KUIJPER A., BRAUN A.: Attack detection in an autonomous entrance system using optical flow. In Crime Detection and Prevention (ICDP 2016), 7rd International Conference on (2016), IET, pp. 1-6.
4. SIEGMUND D., HANDTKE D., KAEHM O.: Verifying isolation in a mantrap portal via thermal imaging. In 2016 International Conference on Systems, Signals and Image Processing (IWSSIP) (May 2016), pp. 1-4.
5. SIEGMUND D., KAEHM O., HANDTKE D.: Rapid classification of textile fabrics arranged in piles. In Proceedings of the 13th International Joint Conference on e-Business and Telecommunications (2016), pp. 99-105.
6. SIEGMUND D., PRAJAPATI A., KIRCHBUCHNER F., KUIJPER A.: An integrated deep neural network for defect detection in dynamic textile textures. In International Workshop on Artificial Intelligence and Pattern Recognition (2018), Springer, pp. 77-84.
7. SIEGMUND D., SAMARTZIDIS T., FU B., BRAUN A., KUIJPER A.: Fiber defect detection of inhomogeneous voluminous textiles. In Mexican Conference on Pattern Recognition (2017), Springer, pp. 278-287.
8. SAMATZIDIS T., SIEGMUND D., GOEDDE M., DAMER N., BRAUN A., KUIJPER A.: The dark side of the face: Exploring the ultraviolet spectrum for face biometrics. In 2018 International Conference on Biometrics (ICB) (Feb 2018), pp. 182-189.
9. SIEGMUND D., WAINAKH A., BRAUN A.: Verification of single-person access in a mantrap portal using rgb-d images. In Proceedings of XII Workshop de Visao Computacional (WVC) (pp. 177-182) (Nov 2016)
10. GOEDDE, M., GABLER, F., SIEGMUND, D., & BRAUN, A. (2018, July). Cinematic narration in VR-Rethinking Film conventions for 360 degrees. In International Conference on Virtual, Augmented and Mixed Reality (pp. 184-201). Springer, Cham.
11. SIEGMUND, D., SAMARTZIDIS, T., Damer, N., NOUAK, A., & BUSCH, C. (2014). Virtual Fitting Pipeline: Body Dimension Recognition, Cloth Modeling, and On-Body Simulation. VRIPHYS, 14, 99-107.

12. SIEGMUND, D., CHIESA, L., HÖRR, O., GABLER, F., BRAUN, A., & KUIJPER, A. (2017, July). Talis—A design study for a wearable device to assist people with depression. In 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC) (Vol. 2, pp. 543-548). IEEE.
13. SIEGMUND, D., TRAN, V. P., VON WILMSDORFF, J., KIRCHBUCHNER, F., & KUIJPER, A. (2019, October). Piggybacking Detection Based on Coupled Body-Feet Recognition at Entrance Control. In Iberoamerican Congress on Pattern Recognition (pp. 780-789). Springer, Cham.
14. SCHELLER, D., BAUER, B., KRAJEWSKI, A., COENEN, C., SIEGMUND, D., & BRAUN, A. (2018, July). An Intuitive and Personal Projection Interface for Enhanced Self-management. In International Conference on Distributed, Ambient, and Pervasive Interactions (pp. 202-213). Springer, Cham.
15. SIEGMUND, D., WAINAKH, A., EBERT, T., BRAUN, A., & KUIJPER, A. (2017, November). Text Localization in Born-Digital Images of Advertisements. In Iberoamerican Congress on Pattern Recognition (pp. 627-634). Springer, Cham.
16. SIEGMUND, D., DEV, S., FU, B., SCHELLER, D., & BRAUN, A. (2018, July). A look at feet: recognizing tailgating via capacitive sensing. In International Conference on Distributed, Ambient, and Pervasive Interactions (pp. 139-151). Springer, Cham.
17. SIEGMUND D., FU B., JOSE-GARIA A., SALAHUDDIN A., KIRCHBUCHNER, F., & KUIJPER, A (2020). Study and Research on Detection of Fiber Defects using Keypoints and Deep Learning. In International Journal of Pattern Recognition and Artificial Intelligence IJPRAI 35.05 (2021): 2150016,
18. SIEGMUND, D., SACCO, L.R. and KUIJPER, A., (2020), September. Issue Based OCR Error Prediction in Video Streams. In 2020 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA) (pp. 75-80). IEEE.
19. SIEGMUND, D., KERCKHOFF, F., MAGDALENO, J.Y., JANSEN, N., KIRCHBUCHNER, F. and KUIJPER, A., (2020, September). Face presentation attack detection in ultraviolet spectrum via local and global features. In 2020 International Conference of the Biometrics Special Interest Group (BIOSIG) (pp. 1-5). IEEE.

A.2. Patents

1. Schmidt, Uwe and Dirk Siegmund.: "Verfahren, Vorrichtung und Computerprogramm zur visuellen Qualitätskontrolle von Textilien." EP2016P58153 (2018).
2. Yeste Magdaleno, Javier and Dirk Siegmund.: "Verfahren zur Klassifizierung von Textilien im freien Fall." EP20186306.5 (2020).

A.3. Talks

1. SIEGMUND D. One Topic Seminar (OOP) on Biometrics, Topic: 3D Face Reference for Multi-Biometrics in Uncontrolled Environments, Zurich, ENFSI (European Network of Forensic Sciences Institute), 09/2016
2. SIEGMUND D. Lecture: User Interaction, Darmstadt, TU Darmstadt, 12/2017 & 12/2018

B. Supervising Activities

The following list summarizes the student bachelor, diploma and master thesis supervised by the author. The results of these works were partially used as an input into the thesis.

B.1. Diploma and Master Thesis

1. Sudeep Dev Madapur Devraj, Prof. Dr. Ing. Dieter Kraus (supervisior), Dirk Siegmund (supervisor), Separation of Subjects in High-Security Locks by Using Capacitive Sensing, Hochschule Bremen, Masters Thesis, 2017
2. Shashank Singh, Prof. Dr. Ing. Arjan Kuijper (supervisior), Dirk Siegmund (supervisor), Speech Emotion Recognition as a wearable device for depressed people, TU Darmstadt, Masters Thesis, 2018
3. Ahmad Masood Salahuddin, Prof. Dr. Ing. Arjan Kuijper (supervisior), Dirk Siegmund (supervisor), Vision based food recognition, TU Darmstadt, Masters Thesis, 2018
4. Ashok Prajapati, Prof. Dr. Ing. Arjan Kuijper (supervisior), Dirk Siegmund (supervisor), Identifying Cuts and Holes in Fabrics, TU Darmstadt, Masters Thesis, 2018
5. Florian Kerckhoff, Prof. Dr. Elke Hergenroether (supervisior), Dr. Olaf Henniger (supervisor), Dirk Siegmund (supervisor), Face Presentation Attack Detection in the UV Spectra, Hochschule Darmstadt, Masters Thesis, 2019
6. Nils Jansen, Prof. Dr. Ing. Arjan Kuijper (supervisior), Dirk Siegmund (supervisor), Depth from Multi-View Single Image, TU Darmstadt, Masters Thesis, 2020

B.2. Bachelor Thesis

1. Laura Chiesa, Oliver Hoerr, Prof. Andrea Krajewski (supervisior), Prof. Dr. Frank Gabler, Dirk Siegmund (supervisor), Emotional User Interfaces Emotionen als ein- und Ausgabemedium, Hochschule Darmstadt, 2016
2. Luis Sacco, Prof. Arjan Kuijper (supervisior), Dirk Siegmund (supervisor), Predicting OCR errors in natural scene images, TU Darmstadt, 2019
3. Tim Wagner, Prof. Arjan Kuijper (supervisior), Dirk Siegmund (supervisor), Applied Food Recognition for Vision-Based Self-Checkout Systems, TU Darmstadt, 2019

Bibliography

[AC21] AKATI J., CONRAD M.: Anti-tailgating solution using biometric authentication, motion sensors and image recognition. In *2021 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)* (2021), IEEE, pp. 825–830. 20, 23

[AEJ21] ABDULLAKUTTY F., ELYAN E., JOHNSTON P.: A review of state-of-the-art in face presentation attack detection: From early development to advanced deep learning and multi-modal fusion methods. *Information fusion 75* (2021), 55–69. 18

[AG19] ANH N. T. H., GIAO B. C.: An empirical study on fabric defect classification using deep network models. In *International Conference on Future Data and Security Engineering* (2019), Springer, pp. 739–746. 25

[AHP06] AHONEN T., HADID A., PIETIKAINEN M.: Face description with local binary patterns: Application to face recognition. *IEEE transactions on pattern analysis and machine intelligence 28*, 12 (2006), 2037–2041. 35

[ALSC02] ARORA H., LEE S., SRIHARI S. N., CHA S.-H.: Individuality of handwriting. *Journal of forensic science 47*, 4 (2002), 1–17. 7, 19

[AMH09] ABDL B., MOHAMMED K., HASHIM S. Z. M.: Handwriting identification: a direction review. In *Signal and Image Processing Applications (ICSIPA), 2009 IEEE International Conference on* (2009), IEEE, pp. 459–463. 20

[Amr16] AMREHN & PARTNER EDV-SERVICE GMBH: Bioporta, Jul 2016. 20

[ASJT17] AARON S. JACKSON ADRIAN BULAT V. A., TZIMIROPOULOS G.: 3d face reconstruction from a single image, 2017. http://cvl-demos.cs.nott.ac.uk/vrn/. 7

[ASK11] A. SRIKAEW K. ATTAKITMONGCOL P. K., KIDSANG W.: Detection of defect in textile fabrics using optimal gabor wavelet network and two-dimensional pca. In *Advances in Visual Computing*. Springer, 2011, pp. 436–445. 27

[ASVN22] AGARWAL A., SINGH R., VATSA M., NOORE A.: Boosting face presentation attack detection in multi-spectral videos through score fusion of wavelet partition images. *Frontiers in big Data 5* (2022), 836749. 46

[AV07] ARTHUR D., VASSILVITSKII S.: k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms* (2007), Society for Industrial and Applied Mathematics, pp. 1027–1035. 104

[BAO17] BUZA E., AKAGIC A., OMANOVIC S.: Skin detection based on image color segmentation with histogram and k-means clustering. In *2017 10th International Conference on Electrical and Electronics Engineering (ELECO)* (2017), IEEE, pp. 1181–1186. 43

[Bat85] BATCHELOR B. G.: Lighting and viewing techniques. *Automated Visual Inspection* (1985). 29

[BAU15] BRUNTON A., ARIKAN C. A., URBAN P.: Pushing the limits of 3d color printing: Error diffusion with translucent materials. *ACM Transactions on Graphics (TOG) 35*, 1 (2015), 4. 41

[Bax96] BAXTER L. K.: *Capacitive Sensors: Design and Applications.* John Wiley & Sons., 1996. 77, 84

[BBLR05] BERTOZZI M., BROGGI A., LASAGNI A., ROSE M.: Infrared stereo vision-based pedestrian detection. In *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE* (2005), IEEE, pp. 24–29. 22

[BDKK22] BOUTROS F., DAMER N., KIRCHBUCHNER F., KUIJPER A.: Self-restrained triplet loss for accurate masked face recognition. *Pattern Recognition 124* (2022), 108473. 15

[BDR*20] BOUTROS F., DAMER N., RAJA K., RAMACHANDRA R., KIRCHBUCHNER F., KUIJPER A.: Iris and periocular biometrics for head mounted displays: Segmentation, recognition, and synthetic data generation. *Image and Vision Computing 104* (2020), 104007. 15

[BF15] BORGHESE N. A., FOMASI M.: Automatic defect classification on a production line. *Intelligent Industrial Systems 1*, 4 (2015), 373–393. 27, 28, 91

[BH16] BOURLAI T., HORNAK L.: Face recognition outside the visible spectrum. 16

[BHW11] BRAUN A., HEGGEN H., WICHERT R.: Capfloor–a flexible capacitive indoor localization system. In *Evaluating AAL Systems Through Competitive Benchmarking. Indoor Localization and Tracking*. Springer, 2011, pp. 26–35. 24

[BKR*10] BOURLAI T., KALKA N., ROSS A., CUKIC B., HORNAK L.: Cross-spectral face verification in the short wave infrared (swir) band. In *Pattern Recognition (ICPR), 2010 20th International Conference on* (2010), IEEE, pp. 1343–1347. 16

[Boo16] BOON EDAM: Mantrap portal solution, July 2016. 20

[Bou16] BOURLAI T.: *Face Recognition Across the Imaging Spectrum*. Springer, 2016. 16

[BS89] BOZINOVIC R. M., SRIHARI S. N.: Off-line cursive script word recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 11*, 1 (1989), 68–83. 50, 53, 125

[BTVG06] BAY H., TUYTELAARS T., VAN GOOL L.: Surf: Speeded up robust features. In *Computer vision–ECCV 2006*. Springer, 2006, pp. 404–417. 35, 36, 51, 100, 107, 119

[BWKF15] BRAUN A., WICHERT R., KUIJPER A., FELLNER D. W.: Capacitive proximity sensing in smart environments. *Journal of Ambient Intelligence and Smart Environments 7*, 4 (2015), 483–510. 24

[BZD*15] BEVERIDGE J. R., ZHANG H., DRAPER B. A., FLYNN P. J., FENG Z., HUBER P., KITTLER J., HUANG Z., LI S., LI Y., ET AL.: Report on the fg 2015 video person recognition evaluation. In *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on* (2015), vol. 1, IEEE, pp. 1–8. 5

[CA13] COOKSEY C. C., ALLEN D. W.: Reflectance measurements of human skin from the ultraviolet to the shortwave infrared (250 nm to 2500 nm). In *Proc. SPIE* (2013), vol. 8734, p. 87340N. 16

[CCNA13] CIRIC I., COJBASIC Z., NIKOLIC V., ANTIC D.: Computationally intelligent system for thermal vision people detection and tracking in robotic applications. In *Telecommunication in Modern Satellite, Cable and Broadcasting Services (TELSIKS), 2013 11th International Conference on* (2013), vol. 2, IEEE, pp. 587–590. 23

[Chu17] CHUI M.: Artificial intelligence the next digital frontier. *McKinsey and Company Global Institute 47*, 3.6 (2017). 1

[CMS14] CERMENO E., MALLOR S., SIGUENZA J. A.: Offline handwriting segmentation for writer identification. In *Biometrics and Security Technologies (ISBAST), 2014 International Symposium on* (2014), IEEE, pp. 13–17. 19

[Cor16] CORP M.: Kinect 2.0, 6 2016. 62

[Cou00] COUNCIL H. K. P.: *Textile Handbook 2000*. The Hong Kong Cotton Spinners Association, 2000. 10, 25

[CSB20] CHAMBINO L. L., SILVA J. S., BERNARDINO A.: Multispectral facial recognition: A review. *IEEE Access 8* (2020), 207871–207883. 16

[CTC*19] CHEH C., THAKORE U., CHEN B., TEMPLE W. G., SANDERS W. H.: Leveraging physical access logs to identify tailgating: Limitations and solutions. In *2019 15th European Dependable Computing Conference (EDCC)* (2019), IEEE, pp. 127–132. 20, 21

[CYS12] CHAN T. W., YAP V. V., SOH C. S.: Embedded based tailgating/piggybacking detection security system. In *2012 IEEE Colloquium on Humanities, Science and Engineering (CHUSER)* (2012), IEEE, pp. 277–282. 23

[Dan20] DANTURTHI R. S.: *Security Engineering*. Apress, Berkeley, CA, 2020, pp. 109–159. 20

[DD16] DAMER N., DIMITROV K.: Practical view on face presentation attack detection. In *BMVC* (2016). 18

[DGXZ19] DENG J., GUO J., XUE N., ZAFEIRIOU S.: Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2019), pp. 4690–4699. 16

[DKS*22] DEVI R. M., KEERTHIKA P., SURESH P., SARANGI P. P., SANGEETHA M., SAGANA C., DEVENDRAN K.: Chapter 5 - retina biometrics for personal authentication. In *Machine Learning for Biometrics*, Sarangi P. P., Panda M., Mishra S., Mishra B. S. P., Majhi B., (Eds.), Cognitive Data Science in Sustainable Computing. Academic Press, 2022, pp. 87–104. 38

[DMNRD12] DE MARSICO M., NAPPI M., RICCIO D., DUGELAY J.-L.: Moving face spoofing detection via 3d projective invariants. In *2012 5th IAPR International Conference on Biometrics (ICB)* (2012), IEEE, pp. 73–78. 18

[DO14] DAMER N., OPEL A.: *Multi-biometric Score-Level Fusion and the Integration of the Neighbors Distance Ratio*. Springer International Publishing, Cham, 2014, pp. 85–93. 74

[DON13] DAMER N., OPEL A., NOUAK A.: Performance anchored score normalization for multi-biometric fusion. In *International Symposium on Visual Computing* (2013), Springer, pp. 68–75. 37

[DON14a] DAMER N., OPEL A., NOUAK A.: Biometric source weighting in multi-biometric fusion: Towards a generalized and robust solution. In *22nd European Signal Processing Conference, EUSIPCO 2014, Lisbon, Portugal, September 1-5, 2014* (2014), pp. 1382–1386. 20, 51

[DON14b] DAMER N., OPEL A., NOUAK A.: Biometric source weighting in multi-biometric fusion: Towards a generalized and robust solution. In *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European* (2014), IEEE, pp. 1382–1386. 37

[DT16] DING C., TAO D.: A comprehensive survey on pose-invariant face recognition. *ACM Transactions on intelligent systems and technology (TIST) 7*, 3 (2016), 37. 6

[DTS06] DALAL N., TRIGGS B., SCHMID C.: *Human Detection Using Oriented Histograms of Flow and Appearance*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 428–441. 127

[dZ06] DE ZEEUW F.: Slant correction using histograms. *Undergraduate Thesis. http://www. ai. rug. nl/˜ axel/teaching/bachelorprojects/zeeuw_slant correction. pdf* (2006). 50

[EEVG*15] EVERINGHAM M., ESLAMI S. A., VAN GOOL L., WILLIAMS C. K., WINN J., ZISSERMAN A.: The pascal visual object classes challenge: A retrospective. *International journal of computer vision 111*, 1 (2015), 98–136. 118

[EKXS96] ESTER M., KRIEGEL H.-P., XU X., SANDER J.: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (1996), AAAI Press, pp. 226–231. 112

[EN101] DIN EN: *Glass in building - Security glazing - Testing and classification of resistance against bullet attack*, 2000-01. DIN EN 1063. 56

[EN109] DIN EN: *Pedestrian doorsets, windows, curtain walling, grilles and shutters - Burglar resistance - Requirements and classification*, 2011-09. DIN EN 1627. 56

[FAKD22] FANG M., ALI H., KUIJPER A., DAMER N.: Patchswap: Boosting the generalizability of face presentation attack detection by identity-aware patch swapping. In *2022 IEEE International Joint Conference on Biometrics (IJCB)* (2022), pp. 1–10. 15

[Far03] FARNEBÄCK G.: Two-frame motion estimation based on polynomial expansion. In *Proceedings of the 13th Scandinavian Conference on Image Analysis* (Berlin, Heidelberg, 2003), SCIA'03, Springer-Verlag, pp. 363–370. 73

[Fit88] FITZPATRICK TB: The validity and practicality of sun-reactive skin types i through vi. *Archives of Dermatology 124*, 6 (1988), 869–871. 32, 34, 41

[FMPZD18] FONNEGRA R. D., MOLINA A., PÉREZ-ZAPATA A. F., DÍAZ G. M.: Mspecface: A dataset for facial recognition in the visible, ultra violet and infrared spectra. In *Technology Trends: Third International Conference, CITT 2017, Babahoyo, Ecuador, November 8-10, 2017, Proceedings 3* (2018), Springer, pp. 160–170. 16, 17

[FS16] FLIR SYSTEMS I.: Flir ax8 camera, Feb. 2016. 57

[FSK02] F.H. SHE L.X. KONG S., KOUZANI A.: Intelligent animal fiber classification with artificial neural networks. *Textile research journal 72*, 7 (2002), 594–600. 27

[Ful97] FULTON J. E.: Utilizing the ultraviolet (uv detect) camera to enhance the appearance of photodamage and other skin conditions. *Dermatologic Surgery 23*, 3 (1997), 163–169. 16

[FYGW15] FENG G., YANG Y., GUO X., WANG G.: A smart fiber floor for indoor target localization. *IEEE Pervasive Computing 14*, 2 (2015), 52–59. 24

[GBDR*22] GOMEZ-BARRERO M., DROZDOWSKI P., RATHGEB C., PATINO J., TODISCO M., NAUTSCH A., DAMER N., PRIESNITZ J., EVANS N., BUSCH C.: Biometrics in the era of covid-19: challenges and opportunities. *IEEE Transactions on Technology and Society* (2022). 15

[Gir15] GIRSHICK R.: Fast r-cnn. *arXiv preprint arXiv:1504.08083* (2015). 117

[GKWS09] GEHRIG D., KUEHNE H., WOERNER A., SCHULTZ T.: Hmm-based human motion recognition with optical flow data. In *2009 9th IEEE-RAS International Conference on Humanoid Robots* (Dec 2009), pp. 425–430. 127

[GS99] GEVERS T., SMEULDERS A. W.: Color-based object recognition. *Pattern recognition 32*, 3 (1999), 453–464. 104

[GS03] GEVERS T., STOKMAN H.: Classifying color edges in video into shadow-geometry, highlight, or material transitions. *Multimedia, IEEE Transactions on 5*, 2 (2003), 237–243. 28

[GZ19] GUO G., ZHANG N.: A survey on deep learning based face recognition. *Computer vision and image understanding 189* (2019), 102805. 16

[HCK*17] HSIAO R.-S., CHEN T.-X., KAO C.-H., LIN H.-P., LIN D.-B.: An intelligent access control system based on passive radio-frequency identification. *Sensors and Materials 29*, 4 (2017), 355–362. 21

[HI06] HOSHINO T., IZUMI T.: Improvement of head extraction for height measurement by combination of sphere matching and optical flow. In *SICE-ICASE, 2006. International Joint Conference* (2006), IEEE, pp. 1607–1612. 22

[Hir05] HIRSCHMÜLLER H.: Accurate and efficient stereo processing by semi-global matching and mutual information. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* (2005), vol. 2, IEEE, pp. 807–814. 104

[HMOS12] HO T. K., MATTHEWS K., O'GORMAN L., STECK H.: Public space behavior modeling with video and sensor analytics. *Bell Labs Technical Journal 16*, 4 (2012), 203–217. 21

[HNN05] H.Y.T. NGAN G.K.H. PANG S. Y., NG M.: Wavelet based methods on patterned fabric defect detection. *Pattern recognition 38*, 4 (2005), 559–576. 28

[HOGF*19] HERNANDEZ-ORTEGA J., GALBALLY J., FIERREZ J., HARAKSIM R., BESLAY L.: Faceqnet: Quality assessment for face recognition based on deep learning. In *2019 International Conference on Biometrics (ICB)* (2019), IEEE, pp. 1–8. 16

[HPS06] HEIKKILÄ M., PIETIKÄINEN M., SCHMID C.: Description of interest regions with center-symmetric local binary patterns. In *ICVGIP* (2006), vol. 6, Springer, pp. 58–69. 36

[HPS09] HEIKKILÄ M., PIETIKÄINEN M., SCHMID C.: Description of interest regions with local binary patterns. *Pattern recognition 42*, 3 (2009), 425–436. 35

[HRBLM07] HUANG G. B., RAMESH M., BERG T., LEARNED-MILLER E.: *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*. Tech. Rep. 07-49, University of Massachusetts, Amherst, October 2007. 35

[HS20] HE S., SCHOMAKER L.: Fragnet: Writer identification using deep fragment networks. *IEEE Transactions on Information Forensics and Security 15* (2020), 3013–3022. 5, 19

[HSBAF22] HAGSTRÖM A. L., STANIKZAI R., BIGUN J., ALONSO-FERNANDEZ F.: Writer recognition using off-line handwritten single block characters. *arXiv preprint arXiv:2201.10665* (2022). 31

[HSC*22] HUANG S., SONG G., CHEN W., QIN J., LIU X., ZHANG B., ZHANG Z.: Time series–based detection on tailgating fare evasions using human pose estimation. *Journal of Transportation Engineering, Part A: Systems 148*, 7 (2022), 04022035. 20

[HWS15] HE S., WIERING M., SCHOMAKER L.: Junction detection in handwritten documents and its application to writer identification. *Pattern Recognition 48*, 12 (2015), 4036–4048. 19

[HZRS16] HE K., ZHANG X., REN S., SUN J.: Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 770–778. 117

[IAMA06] ISLAM M. A., AKHTER S., MURSALIN T. E., AMIN M. A.: A suitable neural network to detect textile defects. In *International Conference on Neural Information Processing* (2006), Springer, pp. 430–438. 28

[Int16] INTERNATIONAL STANDARDS ORGANIZATION: Iso/iec 30107-1:2016 - information technology – biometric presentation attack detection – part 1: Framework, 2016. 6

[IS15] IOFFE S., SZEGEDY C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167* (2015). 87, 115

[ISO05] ISO/IEC: Iso/iec 19794-5:2005 information technology – biometric data interchange formats – part 5: Face image data., 2005. 5

[JD11] JAIN R., DOERMANN D.: Offline writer identification using k-adjacent segments. In *2011 International Conference on Document Analysis and Recognition* (2011), IEEE, pp. 769–773. 19

[JD14] JAIN R., DOERMANN D.: Combining local features for offline writer identification. In *2014 14th International Conference on Frontiers in Handwriting Recognition (ICFHR)* (2014), IEEE, pp. 583–588. 51

[JGGF16] JOSÉ-GARCÍA A., GÓMEZ-FLORES W.: Automatic Clustering Using Nature-Inspired Metaheuristics: A Survey. *Applied Soft Computing 41* (2016), 192–213. 112

[JJ20] JAVIDI M., JAMPOUR M.: A deep learning framework for text-independent writer identification. *Engineering Applications of Artificial Intelligence 95* (2020), 103912. 19

[JN21] JIN R., NIU Q.: Automatic fabric defect detection based on an improved yolov5. *Mathematical Problems in Engineering 2021* (2021), 1–13. 27

[JSD*14] JIA Y., SHELHAMER E., DONAHUE J., KARAYEV S., LONG J., GIRSHICK R., GUADARRAMA S., DARRELL T.: Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia* (2014), ACM, pp. 675–678. 114

[JSR94] JI T. L., SUNDARESHAN M. K., ROEHRIG H.: Adaptive image contrast enhancement based on human visual properties. *IEEE Transactions on Medical Imaging 13*, 4 (Dec 1994), 573–586. 36

[JWZ*21] JUN X., WANG J., ZHOU J., MENG S., PAN R., GAO W.: Fabric defect detection based on a deep convolutional neural network using a two-stage strategy. *Textile Research Journal 91*, 1-2 (2021), 130–142. 26

[KA15] KINGA D., ADAM J. B.: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)* (2015), vol. 5. 116

[KBG17] KAMBI BELI I. L., GUO C.: Enhancing face identification using local binary patterns and k-nearest neighbors. *Journal of Imaging 3*, 3 (2017), 37. 16

[KFDS13] KLEBER F., FIEL S., DIEM M., SABLATNIG R.: Cvl-database: An off-line database for writer retrieval, writer identification and word spotting. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on* (2013), IEEE, pp. 560–564. 19

[KFS18] KEGLEVIC M., FIEL S., SABLATNIG R.: Learning features for writer retrieval and identification using triplet cnns. In *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)* (2018), IEEE, pp. 211–216. 19

[KJAFA20] KORTLI Y., JRIDI M., AL FALOU A., ATRI M.: Face recognition systems: A survey. *Sensors 20*, 2 (2020), 342. 16

[KK02] KANG T. J., KIM S.: Objective evaluation of the trash and color of raw cotton by image processing and neural network. *Textile Research Journal 72*, 9 (2002), 776–782. 27

[KL03] KUO C.-F. J., LEE C.-J.: A back-propagation neural network for recognizing fabric defects. *Textile Research Journal 73*, 2 (2003), 147–151. 27

[Kla21] KLAESS J.: Why computer vision in manufacturing works best as part of an industry 4.0 ecosystem. *Tulip* (2021). 1

[KSH12] KRIZHEVSKY A., SUTSKEVER I., HINTON G. E.: Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (2012), pp. 1097–1105. 117

[Kum08] KUMAR A.: Computer-vision-based fabric defect detection: a survey. *Industrial Electronics, IEEE Transactions on 55*, 1 (2008), 348–363. 25, 28, 29

[LBBH98] LECUN Y., BOTTOU L., BENGIO Y., HAFFNER P.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE 86*, 11 (1998), 2278–2324. 114

[Lev14] LEVERITT T.: How the sun sees you, 2014. http://s.fhg.de/Qgx. 8

[LGSP13] LOULOUDIS G., GATOS B., STAMATOPOULOS N., PAPANDREOU A.: Icdar 2013 competition on writer identification. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on* (Aug 2013), pp. 1397–1401. 19

[LLL*21] LI C., LI J., LI Y., HE L., FU X., CHEN J.: Fabric defect detection in textile manufacturing: a survey of the state of the art. *Security and Communication Networks 2021* (2021), 1–13. 26

[LLYY12] LI Z., LIU G., YANG Y., YOU J.: Scale- and rotation-invariant local binary pattern using scale-adaptive texton and subuniform-based circular shift. *IEEE transactions on image processing 21*, 4 (2012), 2130–40. 106, 119

[LSD15] LONG J., SHELHAMER E., DARRELL T.: Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 3431–3440. 117

[LSHF95] LIKFORMAN-SULEM L., HANIMYAN A., FAURE C.: A hough based algorithm for extracting text lines in handwritten documents. In *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on* (1995), vol. 2, IEEE, pp. 774–777. 49

[LT21] LIU R., TAN W.: Eqface: A simple explicit quality network for face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 1482–1490. 16

[LUNR15] LIU Z., UKIDA H., NIEL K., RAMUHALLI P.: Industrial inspection with open eyes: Advance with machine vision technology. *Integrated Imaging and Vision Techniques for Industrial Inspection: Advances and Applications* (2015), 1–37. 1

[LWZ13] LUO Y., WU C.-M., ZHANG Y.: Facial expression recognition based on fusion feature of pca and lbp with svm. *Optik-International Journal for Light and Electron Optics 124*, 17 (2013), 2767–2770. 101

[LZF*19] LIU L., ZHANG J., FU X., LIU L., HUANG Q.: Unsupervised segmentation and elm for fabric defect image classification. *Multimedia Tools and Applications 78*, 9 (2019), 12421–12449. 25

[Mac67] MACQUEEN J.: Some Methods for Classification and Analysis of Multivariate Observations. In *Fifth Berkeley Symposium on Mathematics, Statistics and Probability* (1967), University of California Press, pp. 91–110. 112

[MAD14] MATTEO F., ANNALISA F., DAVIDE M.: The magic passport. In *IEEE International Joint Conference on Biometrics (IJCB'14)* (2014), pp. 1–7. 17

[MAIS16] MODIRI ASSARI S., IDREES H., SHAH M.: Human re-identification in crowd videos using personal, social and environmental constraints. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14* (2016),

Springer, pp. 119–136. 22

[MB02] MARTI U.-V., BUNKE H.: The iam-database: an english sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition 5*, 1 (2002), 39–46. 19

[MBR04] MURINO V., BICEGO M., ROSSI I. A.: Statistical classification of raw textile defects. In *in ICPR, 2004* (2004), pp. 311–314. 27, 28

[McG13] MCGOVERN M.: Upgrading your entrance lanes, 2013. 20

[Mis15] MISHRA D.: A survey-defect detection and classification for fabric texture defects in textile industry. *International Journal of Computer Science and Information Security 13*, 5 (2015), 48. 25, 28

[MMB01a] MARTI U.-V., MESSERLI R., BUNKE H.: Writer identification using text line based features. In *Document Analysis and Recognition, 2001. Proceedings. Sixth International Conference on* (2001), IEEE, pp. 101–105. 20

[MMB01b] MARTI U.-V., MESSERLI R., BUNKE H.: Writer identification using text line based features. In *Document Analysis and Recognition, 2001. Proceedings. Sixth International Conference on* (2001), IEEE, pp. 101–105. 49

[MS05] MIKOLAJCZYK K., SCHMID C.: A performance evaluation of local descriptors. *IEEE transactions on pattern analysis and machine intelligence 27*, 10 (2005), 1615–1630. 35

[NB15] NARANG N., BOURLAI T.: Face recognition in the swir band when using single sensor multi-wavelength imaging systems. *Image and Vision Computing 33* (2015), 26–43. 16

[NBH15] NARANG N., BOURLAI T., HORNAK L. A.: Can we match ultraviolet face images against their visible counterparts? In *SPIE Defense+ Security* (2015), International Society for Optics and Photonics, pp. 94721Q–94721Q. 16, 33

[NM19] NGUYEN T.-N., MEUNIER J.: Anomaly detection in video sequence with appearance-motion correspondence. In *Proceedings of the IEEE/CVF international conference on computer vision* (2019), pp. 1273–1283. 21, 23

[NNI*19] NGUYEN H. T., NGUYEN C. T., INO T., INDURKHYA B., NAKAGAWA M.: Text-independent writer identification using convolutional neural network. *Pattern Recognition Letters 121* (2019), 104–112. 19

[NPY11] NGAN H. Y., PANG G. K., YUNG N. H.: Automated fabric defect detection—a review. *Image and Vision Computing 29*, 7 (2011), 442 – 458. 25, 27, 28

[OOA*18] ONI D., OJO J., ALABI B., ADEBAYO A., AMORAN A.: Patterned fabric defect detection and classification (fddc) techniques: A review [j]. *International Journal of Scientific & Engineering Research 9*, 2 (2018), 1156–1165. 26

[OPH96] OJALA T., PIETIKÄINEN M., HARWOOD D.: A comparative study of texture measures with classification based on featured distributions. *Pattern recognition 29*, 1 (1996), 51–59. 36

[OPM02] OJALA T., PIETIKAINEN M., MAENPAA T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence 24*, 7 (2002), 971–987. 35

[PA96] PENEV P. S., ATICK J. J.: Local feature analysis: A general statistical theory for object representation. *Network: computation in neural systems 7*, 3 (1996), 477–500. 35

[PGG*21] PERVAIZ M., GHADI Y. Y., GOCHOO M., JALAL A., KAMAL S., KIM D.-S.: A smart surveillance system for people counting and tracking using particle flow and modified som. *Sustainability 13*, 10 (2021), 5367. 22

[PK05] PRIDDY K. L., KELLER P. E.: *Artificial Neural Networks*. SPIE, Bellingham, WA, 2005. 113

[PSK*10] PERŠ J., SULIĆ V., KRISTAN M., PERŠE M., POLANEC K., KOVAČIČ S.: Histograms of optical flow for efficient representation of body motion. *Pattern Recognition Letters 31*, 11 (2010), 1369–1376. 84

[Rau13] RAUTER M.: Reliable human detection and tracking in top-view depth images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2013), pp. 529–534. 23

[RB17] RAMACHANDRA R., BUSCH C.: Presentation attack detection methods for face recognition systems: A comprehensive survey. *ACM Comput. Surv. 50*, 1 (Mar. 2017), 8:1–8:37. 6, 18

[RBAF15] REBHI A., BENMHAMMED I., ABID S., FNAIECH F.: Fabric defect detection using local homogeneity analysis and neural network. *Journal of Photonics 2015* (2015). 27

[RDB21] RATHGEB C., DROZDOWSKI P., BUSCH C.: Detection of makeup presentation attacks based on deep face representations. In *2020 25th International Conference on Pattern Recognition (ICPR)* (2021), IEEE, pp. 3443–3450. 17

[RHGS15] REN S., HE K., GIRSHICK R., SUN J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (2015), pp. 91–99. 116

[RNR19] REHMAN A., NAZ S., RAZZAK M. I.: Writer identification using machine learning approaches: a comprehensive review. *Multimedia Tools and Applications 78*, 8 (2019), 10889–10931. 31

[RRKB11] RUBLEE E., RABAUD V., KONOLIGE K., BRADSKI G. R.: Orb: An efficient alternative to sift or surf. In *ICCV* (2011), vol. 11, Citeseer, p. 2. 43

[RRV*17] RAGHAVENDRA R., RAJA K. B., VENKATESH S., CHEIKH F. A., BUSCH C.: On the vulnerability of extended multispectral face recognition systems towards presentation attacks. In *2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)* (Feb 2017), pp. 1–8. 18

[RZR*20] RASHEED A., ZAFAR B., RASHEED A., ALI N., SAJID M., DAR S. H., HABIB U., SHEHRYAR T., MAHMOOD M. T.: Fabric defect detection using computer vision techniques: a comprehensive review. *Mathematical Problems in Engineering 2020* (2020), 1–24. 26, 28

[SAS*19] SUN S., AKHTAR N., SONG H., ZHANG C., LI J., MIAN A.: Benchmark data and method for real-time people counting in cluttered scenes using depth sensors. *IEEE Transactions on Intelligent Transportation Systems 20*, 10 (2019), 3599–3612. 22

[SB04] SCHOMAKER L., BULACU M.: Automatic writer identification using connected-component contours and edge-based features of uppercase western script. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 26*, 6 (2004), 787–798. 19, 20

[SBK16] SIEGMUND D., BRAUN A., KUIJPER A.: Stereo-image normalization of voluminous objects improves textile defect recognition. *ISVC 2016, Las Vegas, NV (in Press)* (2016). 12, 27, 91, 109, 110, 111, 118, 119

[SCE01] SALVADOR E., CAVALLARO A., EBRAHIMI T.: Shadow identification and classification using invariant color models. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on* (2001), vol. 3, IEEE, pp. 1545–1548.

104

[Sch93] SCHICKTANZ K.: Automatic fault detection possibilities on nonwoven fabrics. *Melliand Textilberichte 74* (1993), 294–295. 27

[SCP22] SRIVASTAVA A., CHANDA S., PAL U.: Exploiting multi-scale fusion, spatial attention and patch interaction techniques for text-independent writer identification. In *Pattern Recognition: 6th Asian Conference, ACPR 2021, Jeju Island, South Korea, November 9–12, 2021, Revised Selected Papers, Part II* (2022), Springer, pp. 203–217. 19

[SDF*18] SIEGMUND D., DEV S., FU B., SCHELLER D., BRAUN A.: A look at feet: recognizing tailgating via capacitive sensing. In *International Conference on Distributed, Ambient, and Pervasive Interactions* (2018), Springer, pp. 139–151. 12, 24, 84, 89

[SED16] SIEGMUND D., EBERT T., DAMER N.: *Combining Low-Level Features of Offline Questionnaires for Handwriting Identification*. Springer International Publishing, Cham, 2016, pp. 46–54. 12

[SFJG*21] SIEGMUND D., FU B., JOSÉ-GARCÍA A., SALAHUDDIN A., KUIJPER A.: Detection of fiber defects using keypoints and deep learning. *International Journal of Pattern Recognition and Artificial Intelligence 35*, 05 (2021), 2150016. 12

[SFN05] SCHWAB J., FIX R., NICHANI S.: System and method for restricting access through a mantrap portal, Nov. 10 2005. US Patent App. 10/908,557. 21

[SFS*16] SIEGMUND D., FU B., SAMARTZIDIS T., WAINAKH A., KUIJPER A., BRAUN A.: Attack detection in an autonomous entrance system using optical flow. In *Crime Detection and Prevention (ICDP 2016), 7rd International Conference on* (2016), IET, pp. 1–6. 12, 83, 84, 89, 128

[SHK16] SIEGMUND D., HANDTKE D., KAEHM O.: Verifying isolation in a mantrap portal via thermal imaging. In *2016 International Conference on Systems, Signals and Image Processing (IWSSIP)* (May 2016), pp. 1–4. 12, 64, 70, 75, 83, 89, 127

[SIC16] SICK AG: Photoelectric proximity sensor w140-2, Feb 2016. 20

[Sie14] SIEGMUND D.: *Prototypical Development of an In-Shop Advertisement System using Body-Dimension Recognition*. Master's thesis, University of Applied Sciences Darmstadt, 03 2014. 12

[SJK19] SHEHZED A., JALAL A., KIM K.: Multi-person tracking in smart surveillance system for crowd counting and normal/abnormal events detection. In *2019 international conference on applied and engineering mathematics (ICAEM)* (2019), IEEE, pp. 163–168. 22

[SKH16] SIEGMUND D., KAEHM O., HANDTKE D.: Rapid classification of textile fabrics arranged in piles. In *Proceedings of the 13th International Joint Conference on e-Business and Telecommunications* (2016), pp. 99–105. 12, 26, 27, 91, 106, 107, 119

[SKJ16] STEINER H., KOLB A., JUNG N.: Reliable face anti-spoofing using multispectral swir imaging. In *2016 International Conference on Biometrics (ICB)* (June 2016), pp. 1–8. 18

[SKM*20] SIEGMUND D., KERCKHOFF F., MAGDALENO J. Y., JANSEN N., KIRCHBUCHNER F., KUIJPER A.: Face presentation attack detection in ultraviolet spectrum via local and global features. In *2020 International Conference of the Biometrics Special Interest Group (BIOSIG)* (2020), IEEE, pp. 1–5. 12

[SL08] STEINHAGE A., LAUTERBACH C.: Monitoring movement behavior by means of a large area proximity sensor array in the floor. In *BMI* (2008), pp. 15–27. 24

[SLJ*15] SZEGEDY C., LIU W., JIA Y., SERMANET P., REED S., ANGUELOV D., ERHAN D., VANHOUCKE V., RABINOVICH A.: Going deeper with convolutions. *Computer Vision and Pattern Recognition (CVPR)* (2015). 87, 115

[Sme16] SMEENK R.: Kinect v1 and kinect v2 fields of view compared, 6 2016. 62

[Smi96] SMITH J. R.: Field mice: extracting hand geometry from electric field measurements, 1996. 77

[SNL*92] SUEN C. Y., NADAL C., LEGAULT R., MAI T. A., LAM L.: Computer recognition of unconstrained handwritten numerals. *Proceedings of the IEEE 80*, 7 (1992), 1162–1180. 19

[Soe16] SOEHNLE INDUSTRIAL SOLUTIONS GMBH: Weighing platform, Jul 2016. 20

[SOPP18] SOUZA L., OLIVEIRA L., PAMPLONA M., PAPA J.: How far did we get in face spoofing detection? *Engineering Applications of Artificial Intelligence 72* (2018), 368–381. 17

[SPKK18] SIEGMUND D., PRAJAPATI A., KIRCHBUCHNER F., KUIJPER A.: An integrated deep neural network for defect detection in dynamic textile textures. In *International Workshop on Artificial Intelligence and Pattern Recognition* (2018), Springer, pp. 77–84. 12, 27, 116

[SSF*17] SIEGMUND D., SAMARTZIDIS T., FU B., BRAUN A., KUIJPER A.: Fiber defect detection of inhomogeneous voluminous textiles. In *Mexican Conference on Pattern Recognition* (2017), Springer, pp. 278–287. 12, 109, 114

[SSG*18] SAMATZIDIS T., SIEGMUND D., GOEDDE M., DAMER N., BRAUN A., KUIJPER A.: The dark side of the face: Exploring the ultraviolet spectrum for face biometrics. In *2018 International Conference on Biometrics (ICB)* (Feb 2018), pp. 182–189. 12, 32, 41, 43

[SSZ01] SCHARSTEIN D., SZELISKI R., ZABIH R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In *Stereo and Multi-Baseline Vision, 2001. (SMBV 2001). Proceedings. IEEE Workshop on* (2001), pp. 131–140. 66

[STvW*19] SIEGMUND D., TRAN V. P., VON WILMSDORFF J., KIRCHBUCHNER F., KUIJPER A.: Piggybacking detection based on coupled body-feet recognition at entrance control. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 24th Iberoamerican Congress, CIARP 2019, Havana, Cuba, October 28-31, 2019, Proceedings 24* (2019), Springer, pp. 780–789. 12

[SV10] SIDDIQI I., VINCENT N.: Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features. *Pattern Recognition 43*, 11 (2010), 3853–3865. 19

[SWB16] SIEGMUND D., WAINAKH A., BRAUN A.: Verification of single-person access in a mantrap portal using rgb-d images. In *XII Workshop de Visao Computacional (WVC)* (Nov 2016). 12, 75, 83, 89, 127

[SZ11] SUN J., ZHOU Z.: Fabric defect detection based on computer vision. In *Artificial Intelligence and Computational Intelligence*. Springer, 2011, pp. 86–91. 27, 28

[SZ14] SIMONYAN K., ZISSERMAN A.: Very deep convolutional networks for large-scale image recognition. *CoRR abs/1409.1556* (2014). 117

[TAS19] TIRALE M. R. N., AGRAWAL V., SUSHIR Y.: Fabric defect classification using modular neural network. 27

[TBD*21] TERHÖRST P., BOLLER A., DAMER N., KIRCHBUCHNER F., KUIJPER A.: Midecon: Unsupervised and accurate fingerprint and minutia quality assessment based on minutia detection confidence. In *2021 IEEE International Joint Conference on Biometrics (IJCB)* (2021), pp. 1–8. 5, 15

[TC17] TISTARELLI M., CHAMPOD C.: *Handbook of biometrics for forensic science*. Springer, 2017. 15

[TCCGRP*18] TOLEDO-CASTRO J., CABALLERO-GIL P., RODRÍGUEZ-PÉREZ N., SANTOS-GONZÁLEZ I., HERNÁNDEZ-GOYA C.: Beacon-based fuzzy indoor tracking at airports. In *Proceedings* (2018), vol. 2, MDPI, p. 1309. 21

[TCD06] TREPTOW A., CIELNIAK G., DUCKETT T.: Real-time people tracking for mobile robots using thermal vision. *Robotics and Autonomous Systems 54*, 9 (2006), 729–739. 23

[Tig19] TIGERFX: Alginat mask, 2019. https://www.youtube.com/watch?v=qg4UxaazkjA. 41

[TIH*23] TERHÖRST P., IHLEFELD M., HUBER M., DAMER N., KIRCHBUCHNER F., RAJA K., KUIJPER A.: Qmagface: Simple and accurate quality-aware face recognition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (2023), pp. 3484–3494. 16

[TK09] THEODORIDIS S., KOUTRUMBAS K.: *Pattern Recognition*, fourth ed. Elsevier Inc., 2009. 112

[TLLJ10] TAN X., LI Y., LIU J., JIANG L.: Face liveness detection from a single image with sparse low rank bilinear discriminative model. In *European Conference on Computer Vision* (2010), Springer, pp. 504–517. 31

[TP91] TURK M. A., PENTLAND A. P.: Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on* (1991), IEEE, pp. 586–591. 35

[Tuo19] TUOMOLA T.: Applying computer vision to tailgating detection: Case: Kupittaa sports hall. 20, 23

[TVGK09] TAN G. X., VIARD-GAUDIN C., KOT A. C.: Automatic writer identification framework for online handwritten documents using character prototypes. *pattern recognition 42*, 12 (2009), 3313–3323. 19

[VC22] VISION I. . C. C., (CVPR) P. R. C.: Cvpr 2022, 2022. https://cvpr2022.thecvf.com/. 3

[VdS09] VDS SCHADENVERHÜTUNG GMBH: *VdS Guidelines for Alarm Systems, Biometric recognition procedures - Requirements and Test Methods*. Amsterdamer Str. 172-174, 50735 Köln, Germany, 2009. Vds3112. 57

[VJ01] VIOLA P., JONES M.: Fast and robust classification using asymmetric adaboost and a detector cascade. In *NIPS* (2001), vol. 2001, Vancouver, British Columbia,, Canada, pp. 1311–1318. 80

[VMV09] VALTONEN M., MAENTAUSTA J., VANHALA J.: Tiletrack: Capacitive human tracking using floor tiles. In *Pervasive Computing and Communications, 2009. PerCom 2009. IEEE International Conference on* (2009), IEEE, pp. 1–10. 24, 80, 81

[VOBK11] VAN OOSTERHOUT T., BAKKES S., KRÖSE B.: Head detection in stereo data for people counting and segmentation. In *VISAPP* (2011), pp. 620–625. 22

[WD21] WANG M., DENG W.: Deep face recognition: A survey. *Neurocomputing 429* (2021), 215–244. 5, 16

[WGJ*92] WILKINSON R. A., GEIST J., JANET S., GROTHER P., BURGES C. J., CREECY R., HAMMOND B., HULL J. J., LARSEN N., VOGL T. P., ET AL.: *The first census optical character recognition system conference*, vol. 184. US Department of Commerce, National Institute of Standards and Technology, 1992. 19

[WHL14] WANG D., HUANG Z., LU Y.: Text-independent writer recognition using modified texture and microstructure features. In *Progress in Informatics and Computing (PIC), 2014 International Conference on* (2014), IEEE, pp. 75–78. 7, 19

[WHM11] WOLF L., HASSNER T., MAOZ I.: Face recognition in unconstrained videos with matched background similarity. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on* (2011), IEEE, pp. 529–534. 35

[WHY09] WANG X., HAN T. X., YAN S.: An hog-lbp human detector with partial occlusion handling. In *Computer Vision, 2009 IEEE 12th International Conference on* (2009), IEEE, pp. 32–39. 101

[WTB14] WU X., TANG Y., BU W.: Offline text-independent writer identification based on scale invariant feature transform. *Information Forensics and Security, IEEE Transactions on 9*, 3 (2014), 526–536. 51

[XQ16] XING L., QIAO Y.: Deepwriter: A multi-stream deep cnn for text-independent writer identification. In *2016 15th international conference on frontiers in handwriting recognition (ICFHR)* (2016), IEEE, pp. 584–589. 19

[YJHN07] YANG J., JIANG Y.-G., HAUPTMANN A. G., NGO C.-W.: Evaluating bag-of-visual-words representations in scene classification. In *Proceedings of the international workshop on Workshop on multimedia information retrieval* (2007), ACM, pp. 197–206. 100

[YSL*21] YE M., SHEN J., LIN G., XIANG T., SHAO L., HOI S. C.: Deep learning for person re-identification: A survey and outlook. *IEEE transactions on pattern analysis and machine intelligence 44*, 6 (2021), 2872–2893. 22

[Z*04] ZIVKOVIC Z., ET AL.: Improved adaptive gaussian mixture model for background subtraction. In *ICPR (2)* (2004), Citeseer, pp. 28–31. 86

[ZCC18] ZHENG C., CHAM T.-J., CAI J.: T2net: Synthetic-to-realistic translation for solving single-image depth estimation tasks. In *Proceedings of the European conference on computer vision (ECCV)* (2018), pp. 767–783. 133

[ZCM18] ZHAO C., CHEN Y., MA J.: Fabric defect detection algorithm based on mfs and svm. In *2018 International Conference on Image and Video Processing, and Artificial Intelligence* (2018), vol. 10836, International Society for Optics and Photonics, p. 108360H. 27

[ZHH*20] ZHAO Y., HAO K., HE H., TANG X., WEI B.: A visual long-short-term memory based integrated cnn model for fabric defect image classification. *Neurocomputing 380* (2020), 259–270. 26

[Ziv04] ZIVKOVIC Z.: Improved adaptive gaussian mixture model for background subtraction. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on* (Aug 2004), vol. 2, pp. 28–31 Vol.2. 66

[ZP07] ZHAO G., PIETIKÄINEN M.: Improving rotation invariance of the volume local binary pattern operator. In *MVA* (2007), pp. 327–330. 106, 119

[ZWWY16] ZENG K., WU N., WANG L., YEN K. K.: Local visual feature detection and description for non-rigid 3d objects. *Advances in Image and Video Processing 4*, 2 (2016), 01. 107, 119

[ZZLQ16] ZHANG K., ZHANG Z., LI Z., QIAO Y.: Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters 23*, 10 (2016), 1499–1503. 34, 43