The Dark Side of the Face: Exploring the Ultraviolet Spectrum for Face Biometrics

T. Samartzidis, D. Siegmund, M. Goedde, N. Damer, A. Braun, A. Kuijper
Fraunhofer Institute for Computer Graphics Research (IGD)
Fraunhoferstrasse 5, 64283 Darmstadt, Germany
{timotheos.samartzidis, dirk.siegmund, michael.goedde,
naser.damer, andreas.braun, arjan.kuijper}@igd.fraunhofer.de

Abstract

Facial recognition in the visible spectrum is a widely used application but it is also still a major field of research. In this paper we present melanin face pigmentation (MFP) 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.

1. Introduction

Biometric authentication is increasingly gaining popularity. 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[25], face[26] and handwriting[24]. Modalities are differentiated by their universality, uniqueness, constancy, measurability, performance, acceptability and circumvention.

Especially face recognition systems are increasingly used in our daily life. Inexpensive camera sensors, high accuracy and user-acceptance make face recognition one of the most used kind 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 [14] and stored on a passport is used as a reference and compared with a probe image, captured under controlled conditions at the border crossing. 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 [9].

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 finger-print). These multi-biometric system require more time for capturing and processing and often need more than one sensor. This is inconvenient, especially in scenarios, without cooperation of subjects. In this work we will therefore show an amplification of well performing aspects of traditional face recognition algorithms with simultaneously collected extended information of the face.

Recently, Thomas Leveritt [17] showed that image captures in the ultraviolet (UV) wavelength reveal hidden skin pigmentation. Even people that look like they are not having any freckles show sometimes certain pigments in the UV spectrum (see Figure 1 I-III A/B). Melanin face pigmentation (MFP) is naturally caused and indicates cells that got damaged. So far, it is unknown if this pigmentation can be used as biometric feature for face recognition and/or as modality.

In order to proof our assumption that face recognition can be improved using MFP as additional feature, we present a novel face image database with captures in UV and VL spectrum (see Section 3). This database consists of
91 subjects captured over a period of 6 month showing different expressions and poses. The methodology of our verification methods is presented in Section 4. We describe how we used different image descriptors and fusion in a verification scenario for human face recognition. Our results in Section 5 show if MFP provide valuable distinct information, to enhance the performance of face recognition. 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. Melanin Pigmentation

This pigmentation in human tissue has its origin in combinations of e.g., hemoglobin, melanin and water. Therefore, human skin reflects light as a function of wavelength and is different to most other surfaces. Observed spectral reflectance is the result of a combination of absorption and scattering from the surface and within the tissue. The variation over the spectral range appears because of several dominant absorbers referred to as chromophores. Melanin is mainly responsible for freckles, moles and commonly for protection against UV-radiation. High melanin pigmentation occurs when the skin is damaged and is known as freckles (or ephelides), and can be seen in the visible light spectrum (VIS). Lower concentrated MFP are only visible at lower wavelength (UV-MFP), due to the peak absorption of melanin in the 335nm wavelength [16]. As these skin-cells are permanently damaged, they do not change their position over time, but they do change their melanin concentration and their number might increase. As the face is strongly exposed to UV-light from sunlight, most MFP originates in the face.

2. Related Work

Face recognition algorithms have significantly advanced in the last few years by the emergence of new methods like the Deepface algorithm [26]. Their practical use in unconstrained environments and concerns about spoofing still however 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 not detectable in the visual band. 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 systems processing as the consequent way to achieve even better results in multi-spectral face recognition [19, 5]. The UV spectrum in particular has yet to be examined in the context of face recognition even though several studies mention its potential for this use [3, 4]. The observation of the human skin in the UV spectrum has first been examined in the medical field rather than in the context of biometrics. Cooksey et al. [6] measured the reflectance of human skin in specific spectral bands from 250nm (UV) to 2500nm (IR). Narang et al. [20] 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 damages of the skin which are not noticeable in VIS light spectrum [10]. In UV light, these characteristic pigment changes usually appear as dark spots on the skin. In this paper we explore the benefits to use UV images for face recognition with a focus on the MFP.

3. 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 [27].

3.1. 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.

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 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.

3.2. Difference to other recording setups

In comparison to the data shown in an earlier study of Narang et al. [20], 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 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.

3.3. 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 [27] (see Figure 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) 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.

![Image of face effects](image)

**Figure 3 – Effects of suncream on the UV image.**

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.

### 3.4. 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. [30] which was significantly more capable of aligning the UV images then using Eigenfaces [28]. We were still forced to manually align 25 UV-images to guarantee their meaningful inclusion into the dataset. No VIS images were aligned manually.

### 4. 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. [18] and Penev et al. [23] 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 [13] or the YouTube Faces [29] 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. 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.

### 4.1. 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) [22], SURF (Speed Up Robust Features) [2] and CSLBP (Center Symmetric LBP) [12]. CSLBP and LBP are chosen because they are well known and relevant image descriptors in the filed of face recognition [1]. 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 determinin 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 [21, 11, 2]. In the case of SURF, local contrast enhancement was performed (see Figure 4), using a local histogram mapping function [15]. The ideal combination of parameters was compiled throughout the progress of the work.

#### 4.1.1 LBP Feature Vector

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

#### 4.1.2 CSLBP Feature Vector

The features created by CSLBP have only 16 dimensions and are built as described by Heikkilä et al. [11]. 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 × 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.
4.1.3 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:

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

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 Sub-set 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.

4.2. 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. [7]. 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:

$$
f_{PAN}(s) = \begin{cases} s - \min(S_k) & \text{if } s \leq T(S_k) \\
\frac{s - T(S_k)}{\max(S_k) - T(S_k)} & \text{if } s > T(S_k)
\end{cases}
$$

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 [8]. In this approach the overlap area between impostor $S_I^k$ and genuine $S_G^k$ scores is used together with an overall performance measure, the EER, to construct the OLDW 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( \{ S^+_{k}[S > T] \} \cup \{ S^-_{k}[S < T] \} \right) \times EER_k \tag{3}

To extract each weighting \(\omega_k\) for a set in a fusion process we use the equation as shown below:

\[
\omega_k = \frac{1}{\sum_{k=1}^{N} \frac{1}{OLD_k}} \tag{4}
\]

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

\[
f_{\text{Fusion}}(S_k) = \sum_{k=1}^{N} \omega_k \cdot S_k \tag{5}
\]

4.3. Experiments

As described in Section 4 an one versus all approach is used to test the datasets with respective feature types. Eleven experiments (see Figure 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.

5. Results

We created Receiver-Operating-Characteristic curves in order to show the performance of the different image descriptors (see Figure 5 a-c) and fusion approaches (see Figure 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 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 1 and at a FAR of 0.1 in Table 2:

Table 1 – True Acceptance Rate @FAR of 0.01

<table>
<thead>
<tr>
<th></th>
<th>CSLBP-VIS</th>
<th>CSLBP-UV</th>
<th>LBP-VIS</th>
<th>LBP-UV</th>
<th>SURF-VIS</th>
<th>SURF-UV</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSLBP-VIS</td>
<td>0.81</td>
<td>0.83</td>
<td>-</td>
<td>-</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>CSLBP-UV</td>
<td>0.83</td>
<td>0.72</td>
<td>0.91</td>
<td>0.76</td>
<td>-</td>
<td>0.75</td>
</tr>
<tr>
<td>LBP-VIS</td>
<td>-</td>
<td>0.91</td>
<td>0.89</td>
<td>0.88</td>
<td>-</td>
<td>0.74</td>
</tr>
<tr>
<td>LBP-UV</td>
<td>-</td>
<td>0.76</td>
<td>0.88</td>
<td>0.68</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SURF-VIS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.54</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>SURF-UV</td>
<td>0.89</td>
<td>0.78</td>
<td>0.91</td>
<td>-</td>
<td>0.75</td>
<td>0.69</td>
</tr>
</tbody>
</table>

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 rises from 0.81 to 0.83 as seen in Table 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 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 1). But because of the small size of our dataset, we can not contribute with a skin-type wise statistics of performance.

6. Conclusion

We presented novel biometric face features in the UV spectrum which enhance the verification performance of current systems. Those features can be integrated in state of the art face recognition systems by acquiring images of subjects in the UV wavelengths. The well performing aspects of traditional face recognition algorithms are retained, simultaneously the overall accuracy is amplified by using this extended information of the face. Selectivity of the useful parts of data is enhanced by the application of PAN-minmax and the OLDW on the score level. This selectivity of importance in scores is regarded as generally applicable. It should open the groundwork upon which current algorithms like e.g. OpenFace or similar could improve their performances if they would use additional UV face images without changing the basic way of operation. Through the rise of neural networks and deep learning, face recognition in the UV spectrum could add performance increases beyond of what was shown in this exploration of the topic. This presents itself also as a good opportunity for future research.

Due to privacy reasons, the authors are unable to release the database, however we offer the possibility to run tests

<table>
<thead>
<tr>
<th></th>
<th>CSLBP-VIS</th>
<th>CSLBP-UV</th>
<th>LBP-VIS</th>
<th>LBP-UV</th>
<th>SURF-VIS</th>
<th>SURF-UV</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSLBP-VIS</td>
<td>0.94</td>
<td>0.94</td>
<td>-</td>
<td>-</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>CSLBP-UV</td>
<td>0.84</td>
<td>0.87</td>
<td>0.96</td>
<td>0.89</td>
<td>-</td>
<td>0.88</td>
</tr>
<tr>
<td>LBP-VIS</td>
<td>-</td>
<td>0.96</td>
<td>0.96</td>
<td>0.95</td>
<td>-</td>
<td>0.96</td>
</tr>
<tr>
<td>LBP-UV</td>
<td>-</td>
<td>0.89</td>
<td>0.95</td>
<td>0.83</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SURF-VIS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.79</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>SURF-UV</td>
<td>0.95</td>
<td>0.88</td>
<td>0.96</td>
<td>-</td>
<td>0.90</td>
<td>0.83</td>
</tr>
</tbody>
</table>
Figure 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.
on our data if a ready to run executable is provided. For this purpose the structure of the data will be released.

Acknowledgment

This work was supported by the German Federal Ministry of Education and Research (BMBF) as well as by the Hessen State Ministry for Higher Education, Research and the Arts (HMWK) within CRISP.

References

Citation