

Skin Detail Analysis for Face Recognition

Jean-Sébastien Pierrard, Thomas Vetter
University of Basel, Department of Computer Science
Bernoullistrasse 16, CH - 4056 Basel, Switzerland
`{Jean.Pierrard, Thomas.Vetter}@unibas.ch`

Abstract

This paper presents a novel framework to localize in a photograph prominent irregularities in facial skin, in particular nevi (moles, birthmarks). Their characteristic configuration over a face is used to encode the person's identity independent of pose and illumination. This approach extends conventional recognition methods, which usually disregard such small scale variations and thereby miss potentially highly discriminative features. Our system detects potential nevi with a very sensitive multi scale template matching procedure. The candidate points are filtered according to their discriminative potential, using two complementary methods. One is a novel skin segmentation scheme based on gray scale texture analysis that we developed to perform outlier detection in the face. Unlike most other skin detection/segmentation methods it does not require color input. The second is a local saliency measure to express a point's uniqueness and confidence taking the neighborhood's texture characteristics into account. We experimentally evaluate the suitability of the detected features for identification under different poses and illumination on a subset of the FERET face database.

1. Introduction

Facial skin exhibits various small scale structures in the surface (wrinkles, scars) and the texture (nevi – a general term for pigment lesions like birthmarks and moles) that stand out from normal skin appearance and represent potentially valuable references for individual distinction. Among such skin irregularities moles are especially suited for identification. Their predictable appearance, also under changing illumination, facilitates detection. And their numerous appearance in conjunction with unique distribution patterns scales well with extensive galleries. Furthermore moles require no abstract encoding, in contrast to most other facial features. This fact could be exploited to query a database without having to provide a sample face, *e.g.* “*search all faces with a birthmark near the upper right lip*”.

The goal of this paper is to present techniques for detection and validation of moles that are prominent enough to be used for identification. Relying on such small scale variations is an unusual approach in face recognition. Conventional recognition algorithms are designed to work on low resolution images. For example the well known Eigenfaces approach [12], representative for linear appearance based subspace methods, performs dimensionality reduction using PCA on the raw image data and thereby implicitly treats local variations as noise. Also model based algorithms like the Active Appearance Model in 2D [4] or the Morphable Model in 3D [1] use PCA to model intra class variations. These methods cannot capture small unexpected details in their reconstruction without severe overfitting, which would render the whole method useless. There exist many techniques based on local descriptors using *e.g.* textons, DCT coefficients or Gabor wavelet features. However, none of these methods involve an explicit representation of one of the aforementioned skin features. To our knowledge the only other attempt to exploit mole-like features for identification was presented by Lin and Tang [8]. Their work comprises a multilayer representation of a face in global appearance, facial features (organs), skin texture and irregularities, which all contribute to the identity. The SIFT framework is used for detection and description of irregular skin details which are then combined in an elastic graph for recognition. Their approach also tackles stability and distinctiveness issues by validating interest regions using multiple gallery samples per person and by ensuring dissimilarity to normal skin regions. Unfortunately the authors do not provide methods how to compute the partitioning and correspondence for the local regions (organs and skin).

We introduce a novel framework capable of detecting prominent moles. We attach importance to the fact that, in contrast to [8], our methods are not bound to a certain recognition scenario. That means in particular that we work independent of pose and illumination and that we are able to deduce personally identifying features from one gallery

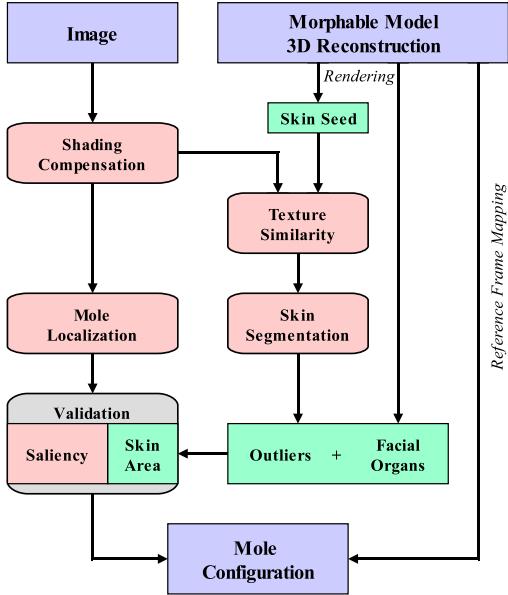


Figure 1. Overview of processing steps and dependencies in our mole detection framework. The left lane shows the main processing steps to obtain locations and saliency measures for moles. Starting from a Morphable Model reconstruction, the right lanes illustrate how the prior knowledge of the 3D face model is incorporated into the system.

image per individual. Our system is divided into three main steps corresponding to the three properties that characterize a local region as birthmark, see also Figure 1.

Appearance: From distance a mole appears simply as small dark region of circular shape surrounded by a brighter area, *i.e.* a so called 'blob'. This description also holds under varying viewing conditions (pose/illumination). We employ a very sensitive multi scale detection scheme, see section 3, to identify even the most subtle mole candidates. **Location:** Due to its sensitivity, the detector also responds to typical facial features such as nostrils, corners of eyes, eyebrows and mouth as well as to unexpected deviations like hair strands. These points are not discriminative across individuals and it is crucial for our scheme that they are rejected. With prior knowledge derived from the Morphable Model we compute in section 4 a binary segmentation of the face in order to rule out points in those non-skin areas. In contrast to most other skin detection/segmentation schemes [13] our approach is texture based and therefore requires no color input or training.

Context: Finally the notion of saliency is introduced in section 5 which allows us to assess the importance of each birthmark candidate for recognition. This procedure takes the relationship between a point's size and contrast and the texture of its neighborhood into account. In essence it represents a combined measure of uniqueness and confidence. Points below a certain saliency threshold are immediately discarded.

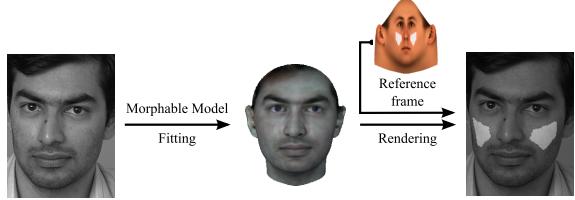


Figure 2. Illustration of the feature/region mapping technique based on dense correspondence. In this example a binary mask, marking the cheeks, was selected in the reference frame. Via a fitting of the Morphable Model this mask can be mapped to the photograph and vice versa detected skin features can be mapped to the model.

2. Face representation

Our framework incorporates face-class specific knowledge from the 3D Morphable Model developed by Blanz and Vetter [1]. The model is derived from 3D laser scans, registered to a common reference in which shape and texture are parameterized. A statistical analysis of this data is performed by Principal Component Analysis. The resulting sets of orthogonal eigenvectors represent facial prototypes which can be linearly combined to form new faces. For a set of model parameters (linear coefficients) and additional parameters for pose and illumination, a realistic face image can be synthesized. Given a photograph of a face, an analysis-by-synthesis approach [2] is used to fit these parameters such that the rendered face model matches the input image. The result is a complete 3D reconstruction of the face where shape and texture in occluded regions (due to pose) have been automatically replenished from the model.

While the model itself delivers features that can be used for recognition [2], we utilize it primarily as a preprocessing to establish a dense correspondence between the pixels of a face photograph and the fixed reference coordinate system in which the model's vertices are parameterized. Thereby we benefit from the model's capabilities in two ways:

- Vertices marking certain features or regions in a face can be consistently selected in the reference frame. Given an image and its Morphable Model reconstruction, these vertices can then be projected into the image domain, either as point sets or as triangular mesh. This technique, shown in Figure 2, allows us to roughly localize the facial organs as well as to determine regions which are likely to contain only skin. The latter will serve our skin segmentation algorithm as initialization.
- With the reference frame acting as intermediary, locations of feature points in different images can be encoded and compared in a pose-independent manner. That means, once we localized interesting moles, their positions within the face are mapped to the reference frame. In these "universal" coordinates we then match the individual mole configurations for recognition.

The optimization algorithm which fits the model to an image [2] requires manually defined feature points like the tip of the nose or corners of the eyes for initialization. Once a 3D reconstruction is computed the corresponding points from the 3D model can be projected back into the image. We compute the average distance between manually defined and reconstructed coordinates to get an indicator for the quality of the fit in terms of correspondence. In the following we shall refer to this measure as *alignment error*. If not stated otherwise we perform all computations only within the support region of the model. This is sufficient since the detected moles must be compared in a common coordinate system which is only defined within the spatial limits of the reconstruction.

3. Mole candidate detection

We detect moles by means of normalized cross correlation (NCC) matching. A Laplacian-of-Gaussian filter mask serves as template, because of its very close resemblance to the blob-like appearance of moles. NCC is not scale invariant and the object size is not known a priori. This requires us to compute the matching for several resolutions, using templates of varying scale. With a growing number of resolutions a straight forward implementation becomes very inefficient. Therefor, inspired by Mikolajczyk *et al.* [9], we carry the matching out in separated steps for candidate point localization in space and scale respectively.

First NCC is computed for a small subset of scales, distributed across the desired search range. In the output image of each scale s_k we then find all local maxima $(x_i, y_i; s_k)$ to pinpoint candidate positions in 2D. Only these points are further considered. In the second step we compute correlation coefficients for the remaining points, using templates that correspond to mole sizes in the range $[.5s_k, 2s_k]$. If the maximum response across these scales is below a fixed threshold the point is discarded. Otherwise the template with maximal correlation defines the points scale for subsequent processing.

Handling scale and space independently has the drawback of causing duplicate point detections, meaning candidates located at different scales and/or coordinates but actually responding to the same feature in the image. We identify such cases and remove all duplicates except for the one with largest scale.

Another problem arises in areas of changing brightness as cause of shading (changing shape or illumination). The intensity gradients surrounding a mole conflict with the uniform area assumption coded in the mole templates. An example for which the described method fails, can be seen in Figure 3. The two obvious solutions to handle such situations are not applicable for us. 1) Lowering the correlation threshold would produce too many false positives in less problematic facial regions. 2) Matching against additional

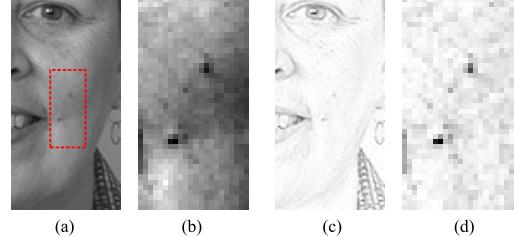


Figure 3. Example of two prominent moles where detection in original image (a) fails. The magnified section (b) shows multiple gradients in vicinity of both moles. After applying our illumination compensation to this region (c,d), the detector succeeds on both moles. Intensities in (b,d) are normalized.

templates on multiple scales that also incorporate skin shading would dramatically increase the computational effort. Instead, we compensate for the shading in the input image. In section 4.2, in the context of skin segmentation, we introduce an image transformation that attempts to remove gradients from shading. The mole detector is then simply applied to the output of this procedure, with better results.

Before conducting the experiments, we hand-labeled in a few gallery images the locations of moles that we deemed salient. The number of scales (range & sample steps) and the NCC threshold were then chosen such that all marked points could be located. Template detection typically reduced the number of candidates for further processing to 1-2% of the pixels representing a face.

4. Facial skin segmentation

The template detector does not incorporate any specific knowledge as to where moles can appear. As consequence it may nominate any facial feature with similar appearance, *e.g.* pupils, nostrils or corners of the mouth. Moreover we must expect sporadic hits in areas with hair (beard, hairstyle). Since none of these findings are characteristic for a person, they have a negative impact on the recognition performance and must therefore be eliminated. We tackle this problem by computing a binary segmentation of the face into skin and “non-skin” regions. Mole candidates lying outside the skin segment can then be rejected. Note that it is not possible to reverse the execution order of detection and then rejection. This becomes clear when we look at the segmentation results at the end of section 4.1.

The non-skin region is composed of two parts. One part is derived directly from the 3D reconstruction with the Morphable Model. In the reference frame of the model we have defined the subset of vertices which belong to the eyes, nostrils and the lips and then projected this selection to the image (see section 2). Due to imperfect reconstructions the resulting mask may not be very precise. We take this into account by dilating the mask according to the measured alignment error.



Figure 4. Display of reconstructed locations for eyebrows, eyes, cheeks and lips on two faces with problematic model fitting. Important features are globally misaligned, due to outliers/hairstyle (left) and expression (right). These shortcomings necessitate an alternative method to mask such regions.

The second non-skin component marks outliers. By this term we refer to all kind of unexpected objects in the sense that they do not appear in every face. Unfortunately the Morphable Model offers no clear strategy how to deal with outliers. On one hand eyebrows, beard and the hair line can be reproduced in the texture. On the other hand hairstyle or open mouth (any expressions) are not represented. Even worse, if larger areas of the face contain outliers this can seriously perturb the model parameters and corrupt a reconstruction in several ways: **1)** Due to the holistic representation, adapting the modeled texture to outliers comes at the cost of higher reconstruction errors in other regions. As result differences between the real and the rendered image “even out”. **2)** The estimated illumination parameters are diverted and can introduce cast shadows into the synthesized image. **3)** The reconstructed shape deteriorates and leads to bad correspondence and thereby misaligned features. Two such problematic examples can be seen in Figure 4. Our conclusion from these shortcomings is that we cannot trust the model’s reconstruction error to reveal all outliers. Instead, we utilize only the most reliable contributions of the model, *i.e.* pose and parts of the shape, to initialize a general purpose segmentation algorithm.

In gray scale skin and non-skin are only distinguishable by their luminosity and/or texture. It is often the case, that their respective intensity distributions show significant overlap, which makes per pixel decisions inappropriate. The techniques presented below operate on local neighborhoods. In the next section we propose a simple procedure to find skin regions by example. Then we introduce a preprocessing step to render the results more robust against varying lighting conditions, see section 4.2. For better understanding we initially motivate these methods using only thresholding as binary segmentation. Later, in section 4.3 we compare our results with a state-of-the-art segmentation algorithm.

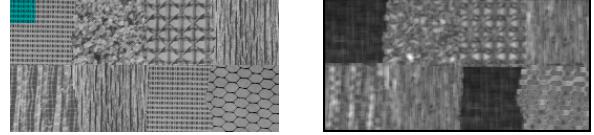


Figure 5. Texture similarity procedure applied to multi-textured image (left), with $I^{tgt} = I^{src}$ and I^{seed} shown in color. The corresponding matching error E_{ts}^k (right) clearly marks regions containing the sample texture.

4.1. Segmentation based on texture similarity

In the work of Efros and Leung [5] textures are synthesized by repeatedly matching the neighborhoods around unprocessed pixels in the synthesis image against all possible source patches extracted from a sample texture. The center pixels of the minimum error patches then build up the synthesized texture. With modifications this idea can be used as analysis tool to compute a measure of texture similarity for one image (*target*) with respect to a given sample of the texture (*source*).

Let I^{tgt} be the target image for which the similarity should be computed. Further we denote with I^{src} a source image and with I^{seed} an associated binary mask, both defining a texture sample region. The similarity is then computed for each pixel $p \in I^{tgt}$ independently by taking its local neighborhood N_p^{tgt} and searching within the seed region of I^{src} for the best matching patch N_q^{src} . Unlike [5] we do not impose different weights on the neighborhood’s pixels. The texture similarity error per target pixel p is:

$$E_{ts}(p) := \min_{q | N_q^{src} \subset (I^{src} \cap I^{seed})} \|N_p^{tgt} - N_q^{src}\|_{SSD} \quad (1)$$

This measurement does not yet take the statistics of the sample texture into account. In order to determine how likely a target pixel may originate from this texture we actually compute the k -nearest-neighbors to N_p^{tgt} . The error E_{ts}^k is then defined, analogous to equation (1), as the average of the corresponding closest-patch distances. Figure 5 demonstrates that ideally this procedure will indeed distinguish between textures.

The method is easily adopted to our segmentation problem. We have chosen the cheeks as facial area which is unlikely to contain outliers. As illustrated in Figure 2, we are able to determine the corresponding region in a novel face from its Morphable Model reconstruction, which provides the skin seed mask I^{seed} . Recalling Figure 4, we also see that this particular mapping is very robust in case of problematic fittings. The algorithm is first applied to the original face image with $I^{tgt} = I^{src}$. Under the assumption that the selected seed contains only skin, the output E_{ts}^k inside this area defines the range of matching errors one can expect for similarly textured regions. The maximum of this range is

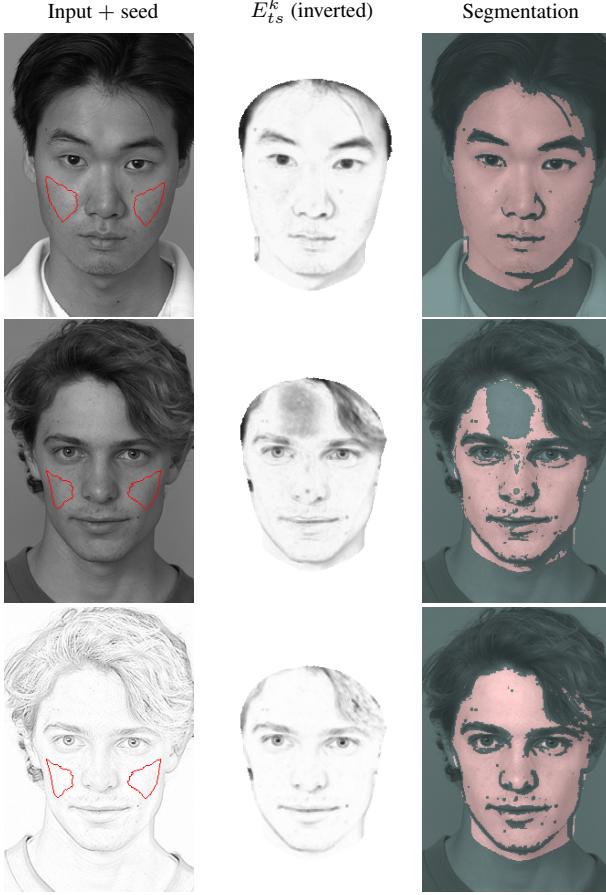


Figure 6. Segmentation by thresholding on the output of texture similarity. In first and second row texture similarity was computed on original image with a very good result in the top row. The bad result in the second row shows how this method can be negatively affected by shading. Performing the same task on an illumination compensated version of the image solved this problem (3rd row).

used as threshold to the entire E_{ts}^k and we obtain a segmentation:

$$I^{skin}(p) := \begin{cases} 1 & \text{if } E_{ts}^k(p) \leq \max_{q \in I^{seed}} E_{ts}^k(q) \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Notice that, without averaging over the k -NN, all errors inside I^{seed} would be zero and this idea could not work. Results for two faces are displayed in Figure 6. We see that shading may provoke large gaps in the segments in lighter areas of the skin. This problem is dealt with in the next section. Another point to notice is that this segmentation method also treats larger moles as outliers. We employ a simple heuristic to prevent such areas from being excluded from further processing: if a mole candidate is located inside a **hole** of the skin segment, we still accept it if the gap's size is less than two times larger than the candidate's scale. Obviously this rule forces us to perform mole detection first.

4.2. Illumination compensation

In 3 and the previous section we pointed out that significant changes in the skin's luminosity can have a negative impact on the performance of the mole detector and the texture similarity algorithm. Therefor we introduce a method to compensate for this effect by performing illumination compensation, based on a variant of homomorphic filtering [6]. The underlying simplified reflectance model assumes that for each pixel location (x, y) the image can be described by the product of reflectance and illumination: $I(x, y) = R(x, y) \cdot L(x, y)$. Thus, to recover R one would simply need to divide the image by the illumination. Unfortunately L is unknown. However, the model further suggests that lighting changes slowly and smoothly across an image while reflectance manifests itself in high frequency components. The idea is now to approximate L by a low-pass filtered version of the image, here denoted by $\mathcal{F}_{lp}(I)$. Since the frequencies of function products are not directly separable (see [6]) this is done in the log-domain. The reflectance becomes:

$$\begin{aligned} \log(R(x, y)) &= \log(I(x, y)) - \log(L(x, y)) \\ &\approx \log(I(x, y)) - [\mathcal{F}_{lp}(\log(I))](x, y) \end{aligned} \quad (3)$$

The exact type and application (spectral or spacial domain) of filter vary among different homomorphic filtering approaches. In our case an approximation to the illumination term is computed by locally fitting smooth functions to the logarithm of image brightness surface.

Given an image I the fitting procedure works as follows. For each pixel p we interpret pixels in its neighborhood N_p as points on a 3D surface. N_p is translated into local coordinates (x_i, y_i, I_i) such that the center pixel p becomes $(0, 0, 0)$. Then we compute a least-squares fit of the quadratic function $z = f(x, y) = \frac{a}{2}x^2 + bxy + \frac{c}{2}y^2$ to these points. Let $\mathbf{z}_p(q)$ denote the LSQ solution for patch N_p evaluated at pixel $q \in N_p$. The approximation induces an error on each pixel of the fitted patch (except on the center pixel). As this procedure is repeated for the whole image, every pixel $p \in I$ receives errors from several patches, namely those neighborhoods which somewhere overlap with p . We accumulate these error contributions separated into positive and negative components:

$$E_{ic}^-(p) := \frac{1}{|N_p|} \sum_{\{q|p \in N_q\}} \min(0, I(p) - \mathbf{z}_q(p)) \quad (4)$$

The definition of E_{ic}^+ is analogous, but rather computes $\max(0, \cdot)$. If this procedure is applied to $\log(I)$, the errors can be interpreted as the right side in equation (3), where the low-pass filter has been implemented as average of smooth function approximations of the neighborhood. Taking the exponential, brings us back to the image domain and results in two reflectance images. The advantage of separating positive and negative errors in equation (4) is that we



Figure 7. Comparison of skin segments obtained from *GrabCut* (bottom) and thresholding (top). *GrabCut* has the advantage of producing fewer small and isolated segments with the downside that it is too conservative in shaded regions.

can isolate different reflectance contributions. For example $R^- = \exp(E_{ic}^-)$ represents the details with darker appearance like creases, moles or pupils, whereas R^+ captures brightness peaks like sharp specular highlights. For the final results in our paper we use only the R^- image as input for the mole detector as well as for the texture similarity algorithm.

4.3. Thresholding vs. “GrabCut”

In order to verify, if we could further improve our skin segmentation, this section compares the results computed by simple thresholding of the texture similarity output E_{ts}^k (obtained on the R^- input) with a modern algorithm based on graph cuts. In graph cut segmentation techniques the pixels of an image are represented as nodes in a graph with the edges reflecting their spatial relationship. The edge weights are defined as some measure of similarity on the connected pixels. Via a cost function, that takes the edges between segments and regional properties into account, the segmentation problem is reformulated into one of partitioning the graph, such that the cost function is minimized.

For the case of two labels (object or background) Boykov and Jolly [3] presented a graph cut formulation that can be efficiently computed with a combinatorial optimization technique and leads to a globally optimal binary labeling. Their cost function is influenced by two models, one for each label, of the pixel values in the respective segment. The models are learned from seed regions in the image, for which the labeling is known. However, with no prior knowl-

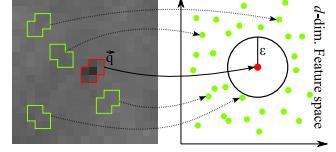


Figure 8. Illustration of density estimation for saliency. A pixel group \vec{q} (red) and all other similarly shaped constellations (green) within its neighborhood populate a multidimensional feature space. The density is estimated by the number of samples lying within a spherical Parzen window around the feature point \vec{q} .

edge on the location of object and background, this placement of seeds demands for user interaction. Rother *et al.* [11] extended this approach. Their method, called *GrabCut* embeds the “one-shot” algorithm of [3] into an EM scheme, where model parameters and labeling are iteratively refined. One major benefit of this procedure is, that it suffices to provide seeds for only one label. Thanks to this property *GrabCut* is directly applicable to our problem setting.

Figure 7 compares results obtained by our own implementation of this algorithm with those from thresholding. In both cases the inputs are the texture similarity error E_{ts}^k computed on the illumination compensated R^- image and the cheek region seed. It is important to note that computations are limited to the support of the Morphable Model reconstruction. This is not relevant for thresholding but for *GrabCut*, because it prevents pixels from the background (clothes, *etc.*) to “pollute” the statistics associated with the two labels. We observe that *GrabCut* (with manually tuned but constant parameters) tends to generate less scattered segments with smoother segment boundaries. However, it cuts off too many pixels in highly shaded regions (especially around the nose) and thus produces larger gaps in the skin segment. This is unacceptable for our current application, since we may lose important moles (also due to the heuristic we implement on segment holes, see section 4.1).

5. Local saliency

Saliency is commonly used as synonym for discriminative power. The more salient a feature is, the better it should be distinguishable from others. The exact definition, however, depends on the actual application. In [14] Walker *et al.* formulate this notion over the probability density in feature space and reason that salient features should lie in low density areas. Hence, intuitively saliency corresponds to rarity. In their paper the pdf is approximated with mixtures of Gaussian kernels. Hall *et al.* [7] take on the same definition but use a more accurate Parzen windows technique for density estimation, which we also adopt here.

Having constrained the detected mole candidates to skin regions, our goal is now to define a measure that allows us to differentiate between prominent and more or less coincidental hits. The latter may occur in “noisy” regions, *e.g.* in

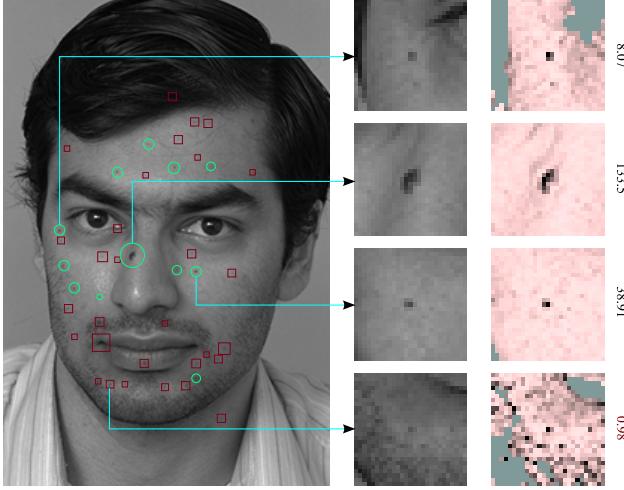


Figure 9. Filtering of mole candidates according to saliency. Circles mark points with saliency $sal_\epsilon \geq 1$ which are later used to identify this face. Zoomed neighborhoods of four candidates with corresponding patches from $\exp(E_{ic}^-)$ show that our saliency measure indeed relates a point's size and contrast to the surrounding noise and delivers an intuitive measure of importance.

the presence of freckles or stubble, where a single dark spot has no significance. A point's scale and correlation coefficient (from detection) contribute to this assessment but are not sufficient. We therefore combine two more properties, the contrast and the uniqueness of a point wrt. its neighborhood, into a saliency value.

Consider a mole candidate composed of a group of d pixels stored as vector \vec{q} . Further assume a square neighborhood $N_{\vec{q}}$ centered around the same location. In our case the width of the neighborhood is fixed and corresponds roughly to the pixel distance between both nostrils in a frontal view. We extract from $N_{\vec{q}}$ all possible translated and mirrored regions \vec{r}_i that have a shape similar to \vec{q} but do not share any pixel with \vec{q} . Let's assume there are M such regions, which populate a d -dimensional feature space. We then consider a hypersphere with radius ϵ and volume V_ϵ around \vec{q} and determine the number k of feature points lying within the sphere, as shown in Figure 8. The ratio $\frac{k/M}{V_\epsilon}$ is then an estimate for the probability density at \vec{q} . Based on the measurement of k we define our saliency as:

$$sal_\epsilon(\vec{q}) := \begin{cases} \min_{\vec{r}_i \subset N_{\vec{q}}, \vec{r}_i \cap \vec{q} = \emptyset} \frac{\|\vec{q} - \vec{r}_i\|}{\epsilon} & \text{for } k = 1 \\ \frac{M-k}{M} & \text{for } k > 1 \end{cases} \quad (5)$$

The radius is chosen as $\epsilon = d \cdot \sigma_{N_{\vec{q}}}^2$, with $\sigma_{N_{\vec{q}}}$ denoting the standard deviation of all pixels in $N_{\vec{q}}$ but not in \vec{q} . Let us take a closer look at the two cases in equation (5):

- $k > 1$: As more points fall within the ϵ -sphere, the estimated density around \vec{q} increases by the ratio $\frac{k}{M}$. The saliency simply decreases by the same rate, taking values in the range $[0, 1)$.

- $k = 1$: No other feature is closer than ϵ to \vec{q} . We compute the distance to its nearest neighbor in multiples of the sphere radius. Since ϵ is related to the sample variance in $N_{\vec{q}}$ the saliency becomes a normalized measure of how much the pixels in \vec{q} stand out from the noise in its neighborhood, ranging from $[1, \infty)$.

We apply the described procedure at every mole candidate location in the illumination compensated image $\exp(E_{ic}^-)$, **constrained to skin segments**. An example of evaluated points is shown in Figure 9. In the left image all points delivered by the mole detection process have been highlighted. The red squares mark candidates which lie either in non-skin regions or which have a computed saliency $sal_\epsilon < 1$. The remaining points are deemed salient and will be used for identification. Of course not all accepted points are equally “interesting”. Figure 9 also depicts the processed patches (right column) of the three most salient moles and one of the rejected points. The non-skin parts are masked out and the remaining pixels are normalized. Clearly the saliency correlates with mole size and is higher for points with less variation (noise) in the surrounding.

6. Experiments

We perform identification purely based on the previously detected moles on a subset (reported in [2]) of the FERET [10] face database, for which we obtained the Morphable Model reconstructions. This subset consists of gray level images with resolutions in the range of 50-80 pixels eye distance. It contains images of 194 individuals in 11 poses from which we chose the sets *ba* (frontal view) for the gallery, *bc-bh* (head rotated by $\pm 40^\circ, \pm 25^\circ, \pm 15^\circ$) and *bk* (frontal view with different illumination) as probe faces. The recognition experiments are limited to persons for which the respective gallery image contains at least one mole with a saliency greater than some threshold. Our similarity measure is based on the mole locations in the Morphable Model reference coordinates and their associated saliency values. We compare two given faces F and G as follows:

- The saliency values of all moles of a face are transformed to relative weights $w_i = sal_i / \sum_{j=1}^n sal_j$.
- A proximity threshold σ_{thr} is defined as the average of the alignment errors of both faces.
- For each mole location i in F we find the closest point j from G . If their distance is smaller than $3\sigma_{thr}$ the point i is considered matched and we define a matching value $v_i = \min(w_i^F, w_j^G) / \max(w_i^F, w_j^G)$. In this case the point j from G is removed so that it cannot match any other locations in F . Otherwise (distance larger than proximity threshold) i remains unmatched, we set $v_i = 0$ and proceed to the next item.

| Saliency threshold (Gallery subset size) | | | | | | |
|--|---------|-------|----------|--------------|---------|--------------|
| | 5 (156) | | 10 (107) | | 15 (83) | |
| Probe | Fail | Perf. | Fail | Perf. | Fail | Perf. |
| <i>bc</i> | 69 | 55.77 | 39 | 63.55 | 26 | 68.67 |
| <i>bd</i> | 34 | 78.20 | 13 | 87.85 | 8 | 90.36 |
| <i>be</i> | 17 | 89.10 | 7 | 93.45 | 4 | 95.18 |
| <i>bf</i> | 20 | 87.18 | 5 | 95.32 | 5 | 93.97 |
| <i>bg</i> | 47 | 69.87 | 24 | 77.57 | 17 | 79.51 |
| <i>bh</i> | 68 | 56.41 | 30 | 71.96 | 21 | 74.70 |
| <i>bk</i> | 42 | 73.07 | 22 | 79.44 | 13 | 84.33 |

Table 1. Performance of identification purely based on detected moles. The gallery (frontal views, *ba*) and probe are limited to faces, which contain at least one mole in the gallery with a saliency greater than the denoted threshold. Performance is listed as number of unidentified faces from the gallery subset (*Fail*) and in percent (*Perf.*).

| Gallery / Probe | Saliency threshold | Gallery subset size | Fail | Perf. |
|-----------------|--------------------|---------------------|------|-------|
| <i>ba / be</i> | 1 | 194 | 42 | 78.35 |
| <i>ba / bf</i> | | | 46 | 76.28 |
| <i>bb / bc</i> | 5 | 180 | 17 | 90.55 |
| <i>bi / bh</i> | 5 | 184 | 21 | 88.58 |

Table 2. Identification from moles on full gallery with $\pm 15^\circ$ rotated probes and on two non-frontal galleries.

- After processing all point from F the matching score is computed as $\sum_{i=1}^{n_F} v_i / \max(n_F, n_G)$.

From the identification results displayed in Table 1 and 2 we notice that: **1)** Performance drops with increasing rotation angle independent of the gallery pose. This is obvious, since the overlapping area in which moles from both faces can be matched shrinks. Recognition under the different illumination (*bk* set) suffers from the lower contrast between moles and skin which results in lower saliency values and thus more rejections. The total number of detected moles is 1600, whereas in all other sets we can account for more than 2200 moles. **2)** At least 80% of the faces have some prominent moles (saliency ≥ 5) for which we obtain recognition performances above 87%. This is quite remarkable, considering that in average about 5-10 locations, representing less than 0.3% of the pixels in a face, determine its identity. Enforcing more prominent moles leads to better performance but greatly reduces the number of usable faces. Somewhere between saliency thresholds of 10-15 is the limit beyond which the number of misclassified faces decreases less than the number of available faces.

7. Conclusion

This paper presented a novel approach to exploit local skin irregularities as features for face identification. We focused on the methodology of detection and evaluation of

such regions and showed that it is possible to determine a persons identity based on only a few well-chosen pixels. Future work comprises refinements in the comparison of local skin features (*e.g.* valuing the absence of salient moles as exclusion criterion) as well as fusion with other face recognition methods to support cases where no moles are present.

References

- [1] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In *Siggraph 1999, Computer Graphics Proceedings*, pages 187–194, Los Angeles, 1999. Addison Wesley Longman.
- [2] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, 2003.
- [3] Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. In *International Conference on Computer Vision, (ICCV)*, pages 105–112, 2001.
- [4] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:681–685, Jan. 2001.
- [5] A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *IEEE International Conference on Computer Vision*, pages 1033–1038, Corfu, Greece, September 1999.
- [6] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2001.
- [7] D. Hall, B. Leibe, and B. Schiele. Saliency of interest points under scale changes. In *British Machine Vision Conference (BMVC'02)*, Cardiff, UK, September 2002.
- [8] D. Lin and X. Tang. Recognize high resolution faces: From macrocosm to microcosm. In *CVPR '06: Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition*, pages 1355–1362, 2006.
- [9] K. Mikolajczyk and C. Schmid. Indexing based on scale invariant interest points. In *Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada*, pages 525–531, 2001.
- [10] P. J. Phillips, H. Wechsler, J. S. Huang, and P. J. Rauss. The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5):295–306, 1998.
- [11] C. Rother, V. Kolmogorov, and A. Blake. “GrabCut”: interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics*, 23(3):309–314, Aug. 2004.
- [12] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.
- [13] V. Vezhnevets, V. Sazonov, and A. Andreeva. A survey on pixel-based skin color detection techniques. In *Proc. Graphicon*, 2003.
- [14] K. Walker, T. Cootes, and C. Taylor. Locating salient object features. In *British Machine Vision Conference (BMVC)*, volume 2, pages 557–566. BMVA Press, 1998.