MODEL-FREE FACE DETECTION AND HEAD TRACKING WITH MORPHOLOGICAL HOLE MAPPING

Udo Ahlvers[†], *Ruben Rajagopalan*[‡], *Udo Zölzer*[†]

[†] Department of Signal Processing and Communications
Helmut-Schmidt-University, Hamburg, Germany
email: `udo.ahlvers@hsu-hh.de`
web: `http://ant.hsu-hh.de`

[‡] Technical University Hamburg-Harburg, Germany
email: `ruben.rajagopalan@tuhh.de`

ABSTRACT

Face detection algorithms have various applications like recognition, expression analysis or video conferencing. Nonetheless, most existing algorithms require additional a-priori knowledge or time-consuming preprocessing steps. In this paper, a model-free feature based approach for face detection and tracking is presented. Possible skin clusters are defined based on one of three different color spaces. In contrast to prior methods, afterwards not the kind or properties of facial characteristics, but only their *existence* is evaluated in a so-called *hole map*. Finally, the detection of faces from these skin clusters is performed with a *mapping process*. This both fast and easy operation can be used for detection and tracking of heads. Good and robust results are obtained for various real-world scenarios.

1. INTRODUCTION

The problem of locating a persons's face in an image or video sequence serves as a first step to higher-level face recognition tasks such as personal identification, security purposes or understanding of facial expressions.
In the recent past there have been extensive surveys and analysis of various face detection algorithms [1, 2], whose conclusions provide promising directions for future research. Generally, all proposed solutions fall under four main categories:

- Knowledge-based
- Feature invariant
- Template matching
- Appearance-based.

The main drawback of most of these algorithms is the necessity of a-priori knowledge (models, rules [3]), user initiation (parameters, templates [4]) or extensive tuning (neural networks [5]). Independently from the chosen approach, a face detection algorithm has to tackle various challenges, e.g. head orientation and motion, occlusion, presence of multiple heads and many more.

In this paper we present a new approach for feature-based face detection, which is model-free and independent from user initiation. Other methods [6, 7] of this class evaluate the kind, position and respective properties of certain facial characteristics (e.g. mouth, nose, eyes or ears) in a partially timeconsuming preprocessing step. Compared to them, here only the **existence** of such features is analysed in a *morphological mapping process*. This makes our algorithm robust and significantly faster, yielding the possibility for real-time applications.

In section 2, an overview of suitable color space skin custers is given. The face detection algorithm with morphological hole mapping is presented in section 3. Results for different real-world scenarios together with a comparison are discussed in section 4. At last, a conclusion and an outlook for future work is given.

2. COLOR SPACE SKIN CLUSTERS

In order to be able to work model-free and non user-initiated, the decision was taken to focus on a feature-based approach. Here color information is employed to detect skin regions. In the recent past, a wide variety of different color spaces has been applied to the problem of skin color modelling [8, 9]. Throuhout this work, three different skin clusters have been implemented. Their declarations and characteristics are given in the following subsections.

2.1 RGB skin cluster

The RGB True Color is an additive color system based on tri-chromatic theory. It is one of the most commonly used color spaces, with a lot of research activities being based on it. The chosen skin cluster for RGB is [10]:

$$\begin{gathered}(R,G,B)\text{ is classified as skin if:}\\ R>95 \text{ and } G>40 \text{ and } B>20\\ \max\{R,G,B\}-\min\{R,G,B\}>15\\ |R-G|>15 \text{ and } R>G \text{ and } R>B\,,\\ \text{where } R,G,B=[0,255]\quad .\end{gathered}$$

2.2 YCbCr skin cluster

In contrast to RGB (see Sec. 2.1), the YCbCr color space is luma-independent, resulting in a better performance. An Example for the YCbCr skin cluster is given in Fig. 1.

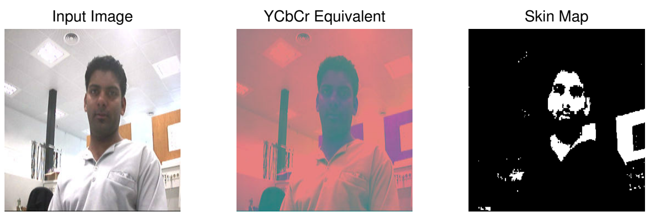

Figure 1: *Illustration of the YCbCr skin map.*

It is extensively used in digital video standards and compression applications. The corresponding skin cluster is given as [11]:

$$
\begin{gathered}
(Y, Cb, Cr) \text{ is classified as skin if:} \\
Y > 80 \\
85 < Cb < 135 \\
135 < Cr < 180\,, \\
\text{where } Y, Cb, Cr = [0, 255] \quad .
\end{gathered}
$$

2.3 HSV skin cluster

The main advantage of this luma-independent color space lies in its simplicity of color specification. Here, the skin cluster is defined by [12]:

$$
\begin{gathered}
(H, S, V) \text{ is classified as skin if:} \\
0 < H < 50 \\
0.23 < S < 0.68\,, \\
\text{where } H = [0, 360] \text{ and } S, V = [0, 1] \quad .
\end{gathered}
$$

Of course, always *only one* color space can be used active at a time for the detection algorithm. In general, it is recommend to use one of the luma-independent skin clusters (e.g. YCbCr or HSV) to extract possible face candidates. This cognition is drawn from the results of other extensive comparative work [8, 9]. Based on own prior experiments, the HSV skin cluster was used for all test scenarios (see. Sec. 4) here. It performs well in most cases with fewer false positives and better skin maps [13].

Up to now, only skin-like regions have been extracted from the image (see Fig. 1). The process of dividing these regions into facial and non-facial areas will be described in the next section.

3. MORPHOLOGICAL FACE EXTRACTION

Numerous methods have been proposed to segment a face from a cluttered background, based on feature information (e.g. [6, 7]). In general, this separation needs a costly preprocessing step to extract and group edges and corners or to locate facial characteristics.

3.1 Generation of Hole Maps

In our approach this process is significantly simplified. First a skin map x_S is calculated based on one of the skin clusters presented in Sec. 2. Then, not the type, but only the *existence* of facial features is evaluated. This is based on the assumption that a major part of holes detected on a skin map x_S will be related to certain facial characteristics (mouth, nose, eyes and ears) [13].
The so-called **hole map** x_H is thus created by

$$x_H(n_1, n_2) = x_{S,C}(n_1, n_2) - x_S(n_1, n_2)\,. \tag{1}$$

In Eq. 1, n_1, n_2 denote the row and column indices of the image to express that this (and all the following) operations have to be carried out for every image pixel. $x_{S,C}$ describes a closed version of the skin map x_S after a dilation and erosion step. Additionally, the so created hole map is cleaned to suppress single noise values. The generation of skin maps $x_S, x_{S,C}$ and hole map x_H from the input image x_I is shown in Fig. 2.

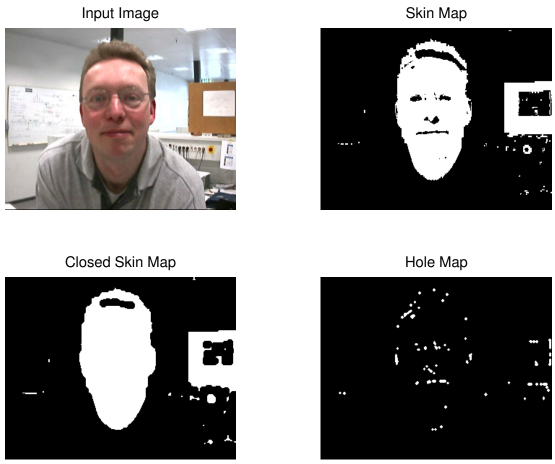

Figure 2: *Illustration of the hole map generation. The hole map is computed according to Eq. 1.*

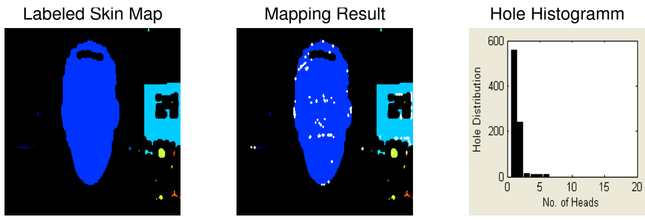

Figure 3: *Hole mapping onto the labeled skin map and corresponding histogram (hole distribution).*

3.2 Generation of Face Maps

In order to detect possible regions for face candidates, the closed skin map is labeled. Here pixels with 4 or 8 connecting neighbours are grouped into one region (label). The number of labels reflects the number of possible face candidates in the image.
Afterwards the hole map (see Sec. 3.1) is **mapped** onto the labeled skin map. A histogram of the hole distribution is calculated which gives the contributions of every region to the hole map. Fig. 3 shows example images for this mapping process.
As certain facial features are responsible for the existence of holes, the label with the maximum hole contribution finally yields the face map. A textured version of the (binary) face map x_F can be easily generated by

$$x_{F,T}(n_1, n_2) = \begin{cases} x_I(n_1, n_2) & \text{if } x_F(n_1, n_2) = 1 \\ 0 & \text{otherwise}\,. \end{cases} \tag{2}$$

Finally, an ellipse is fit over the head. From the face map x_F, the centroid of the face is computed together with its minor and major axes to ensure best fitting. Fig. 4 shows the labeled skin map together with the final face maps. As a final result, the detected face is marked by the ellipse around the head in the input image.
As only the existence of holes is evaluated, this *hole mapping* approach works very well with significantly less computational effort compared to other feature-based face detection algorithms.
The functionality has now to be tested on different real scenes. Results of this test are presented in the next section.

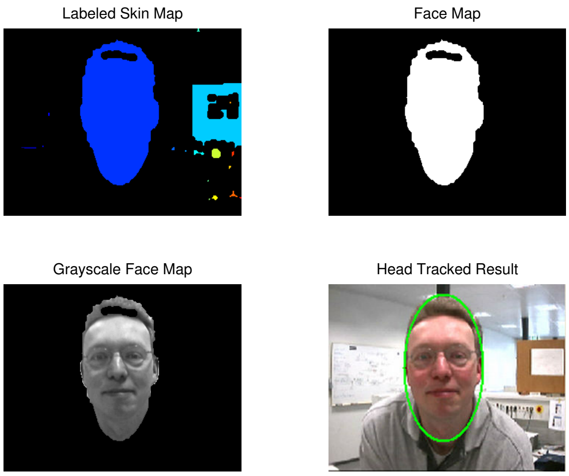

Figure 4: *Face map generation by hole mapping. The textured version (lower left) was computed according to Eq. 2.*

No.	Challenge, Dataset (10 frames length)	Success Rate
1	Presence of Multiple Heads	100
2	Tilting of Heads	100
3	Zooming in and out	100
4	Disappearance & Reappearance	100
5	Occlusion	90
6	360° Rotation	70

Table 1: *Summary of challenges and their success rates. Each video sequence covers a specific task.*

4. RESULTS

Any head tracker has to undergo various challenges while considering real-world scenarios. In our test environment, short sequences were captured consisting of 10 frames (320 x 240 *px*) at approx. 1 frame/second (see Tab. 1). This ensures the existence of motion/changes on the one hand, while on the other hand occuring changes will not be to large [13]. These videos were then processed independently frame-by-frame with the hole-mapping face detection algorithm (see Sec. 3). Here, consciously *no* interframe processing (e.g. kalman filter or motion compensation) was applied in order to focus totally on the face detection approach.
For the evaluation, at first the influence of external conditions was analysed (see Sec. 4.1). Furthermore, a total of 6 different scene contents has been studied during this work. Results for three selected scenarions are presented and discussed in Sec. 4.2 - 4.4. A complete overview together with their respective tracking success rates is given in Tab. 1.

4.1 External Conditions

In a first step, the following *"external"* effects have been evaluated:

- Lighting conditions
- Influence of noise
- Range of displacement from the camera.

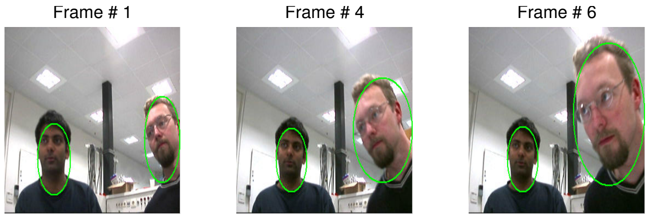

Figure 5: *Presence of multiple heads together with motion (zooming in and out). The ellipse is fit to the changing size of the head.*

We found that our approach shows good performance for varying lighting conditions (see Fig. 5 and 7 for comparison). Especially the results from the luma-independent skin clusters (see Sec. 2.2 and 2.3) were not effected by luminance changes.
Similar results could be observed in the presence of additive noise. This can be caused by low-cost CCD-sensors in the camera (e.g. in webcams) or wrong exposure times during the image capture. As these effects mainly influence the image intensity without chrominance modifications, face detection was possible unaffected.
The applicable range of displacement for successful detection lies between 60 cm and 4 m for the used hardware setup. This can easily be extended by modifying the focal length of the camera lens.

4.2 Presence of Multiple Heads

First, Fig. 5 shows the presence of multiple heads in a scene, while one of them is moving towards the camera (*zooming in*). Based on the hole histogram (see. Fig. 3), the decision is taken how many labeled regions in fact represent a face. With that, the tracker is able to detect and to separate both faces, independently from specific facial features (beard, glasses or face complexion). The position and size of the ellipse is adapted to the moving head very accurate. Same results were achieved for a scenario with tilting motion of the heads.

4.3 Occlusion

Second, the effect of occlusion (disappearance and reappearance of heads) is discussed. As long as at least a part of the face is visible, the tracker follows the face and the ellipse is fit onto the estimated area of the hole face (see Fig. 6). Only in the case of complete occlusion, there is no face detection possible.
After reappearance, the head is detected again. Here, our algorithm does not need any re-initialisation or user input to successfully continue the tracking process! Due to the one missed frame, this can be stated as 90% efficiency (see Tab.1).

4.4 Rotation

Third, results are presented for a 360° rotation of the head (see Fig. 7). Once the decision to work with skin color was taken (see Sec. 2), it was known that the challenge of complete rotation of the head would be difficult for the tracker to handle. The obvious cause is that the contribution of holes from the head would be comparatively low when the head

Figure 6: *Occlusion and reappearance of the head. After reappearance, the tracking continues automatically.*

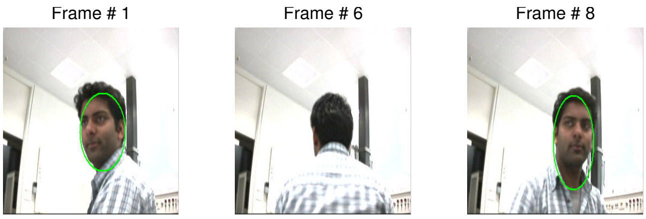

Figure 7: *360° rotation of the head. In case of the back of the head, there are no facial features existent.*

is turned to 180° . As a result, the tracker fails to locate the head in these frames, leading to a success rate of about 70%. When the turning of the head continues, the algorithm starts retracking successfully (reappearance of the face).
There has been an approach proposed [4], where this task has been tackled very well. The drawback there is the need for an user-invented initial state of the head. This is exactly what we wanted to avoid in our approach.

At last, it shall be mentioned that this algorithm works well for different races, too. All these tests were perfomed with people from Europe, Africa and India. The skin clusters proposed in Sec. 2 represented a good base for all test persons, finally leading to the impressive success rates in Tab. 1.

5. CONCLUSION AND OUTLOOK

We have presented a method for model-free, non user-initiated face detection. Based on color space skin clusters, skin-like regions are detected in the image. Afterwards the existence of certain facial characteristics is located in a hole map. In a central step this hole map is then *mapped* onto the labeled skin map. A histogram of the hole distribution gives the contributions of every region to the hole map. In this way it is possible to compute the final face map robustly *without* the need to determine the kind of facial characteristics explicitly.
Applied to several short real-world sequences, the algorithm shows good performance even under various circumstances. Challenges like multiple heads, zooming, tilting or disappearance are handly perfectly. Even in case of failure, subsequent retracking is performed successfully without the need of re-initialisation.

Future work focuses on the integration of an interframe-processing step to handle the problems with 360° rotations. Here the last well-known localisation could be used for a position estimation in uncertain frames. The behaviour on longer test sequences will be studied shortly. Finally, this approach could also be extended to work on higher level face recognition, like expression and gesture analysis.

REFERENCES

[1] D. Yang, N. Ahuja. "Detecting faces in images: A survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 24, no. 1, pp. 34-58, Jan. 2002.

[2] E. Hjelmas, B. Low. "Face detection: A survey", *Computer Vision and Image Understanding*, no. 83, pp. 236-274, 2002.

[3] G. Yang, T. Huang. "Human face detection in complex background", *Pattern Recognition*, vol. 27, pp. 53-63, 1994.

[4] S. Birchfield, "Elliptical head tracking using intensity gradients and color histograms", in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Santa Barbara, USA, June 1998.

[5] H. Rowley, S. Baluja, T. Kanade. "Neural network-based face detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 20, no. 1, pp. 23-38, Jan. 1998.

[6] V. Govindaraju. "Locating human faces in photographs", *Int. Journal of Computer Vision (IJCV)*, vol. 19, no. 2, pp. 129-146, Aug. 1996.

[7] R. Seguier, "A very fast adaptive face detection system", in *Proc. IASTED Conf. on Visualization, Imaging, and Image Processing (VIIP)*, Marbella, Spain, Sep. 2004.

[8] J.C. Terrillon, M. Shirazi, H. Fukamachi, S. Akamatsu, "Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images", in *Proc. IEEE Conf. on Automatic Face and Gesture Recognition (FG)*, Santa Barbara, USA, March 2000, pp. 54-61.

[9] B. Zarit, B. Super, F. Quek, "Comparison of five color models in skin pixel classification", in *Proc. IEEE Int. Workshop on recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems*, Corfu, Greece, Sept. 1999, pp. 58-63.

[10] P. Peer, F. Solina, "An automatic human face detection method", in *Proc. 4th Computer Vision Winter Workshop (CVWW)*, Rastenfeld, Austria, Feb. 1999, pp. 122-130.

[11] G. Kukharev, A. Novosielski, "Visitor identification - elaborating real time face recognition system", in *Proc. 12th Winter School on Computer Graphics (WSCG)*, Plzen, Czech Republic, Feb. 2004, pp. 157-164.

[12] K. Sobottka, I. Pitas, "Segmentation and tracking of faces in color images", in *Proc. IEEE Conf. on Automatic Face and Gesture Recognition (FG)*, Killington, USA, Oct. 1996, pp. 236-241.

[13] R. Rajagopalan, "Face detection and head tracking in real video sequences", Project Work, Helmut-Schmidt-University / Technical University Hamburg-Harburg, Hamburg, Germany, Dec. 2004.