Active Shape Model Based Segmentation and Tracking of Facial Regions in Color Images

Bogdan Kwolek

Rzeszów University of Technology, W. Pola 2, 35-959 Rzeszów, Poland
bkwolek@prz.rzeszow.pl

Abstract. An approach for segmenting and tracking a face in a sequence of color images is presented. It enables reliable segmentation of facial region despite variation of skin-color perceived by a camera. A second order Markov model is utilized to forecast the skin distribution of facial regions in the next frame. The histograms that are constructed from the predicted distribution are backprojected to generate candidates of facial regions. Afterwards, a connected component labeling takes place. Spatial morphological operations, such as size and hole filtering are employed next. The Active Shape Model seeks to match a set of model points to the image. This statistical model of shape supports the segmentation of facial region undergoing tracking. Histograms are accommodated over time using feedback from shape, newly classified skin pixels and predictions of the skin-color evolution. This evolution is described by translation, rotation and scaling. In this context, the novelty of our approach lies in the introduction of Active Shape Model dealing with translation, rotation and scaling of the target to support face verification as well as to guide the evolution of skin distribution. The kernel histograms characterize the face during tracking in subsequent frames. The proposed algorithm achieves reliable detection and tracking results. The resulting system runs in real-time on standard PC computer.

1 Introduction

Skin-color has proven to be effective and robust cue for face detection, localization and tracking. A large part of image processing techniques uses skin detection as a first primitive for subsequent extraction of image features. Skin pixel candidates can be further processed to extract shape or motion cues. Well-known methods of color modeling, such as histograms and Gaussian mixture models have enabled the creation of appropriately exact and fast detectors of skin. Many available vision systems are now applying such techniques to extract skin-color patches for face detection and tracking in video sequences. In particular, skin color based methods are robust to changes in scale, resolution and partial occlusion. However, such techniques are not as good as can be for use in real environments because skin-color perceived by a camera usually changes when the lighting condition varies. Therefore, for reliable detection of skin pixels a dynamic color model that can cope with nonstationary skin-color distribution over time should be applied in vision systems.

The task of finding a human face in an image is referred to as face localization or face segmentation. The grouping of extracted facial features into face candidates, the heuristic rules and knowledge about a typical face and the correlation to statistical face template are examples of approaches commonly employed to detect the face [22]. Two types of information are typically used to perform segmentation during face tracking. The first is color information [4][7][10][14][19]. The second is the geometric configuration of the face shape and even a given set of facial features, e.g. both eyes, the nose, mouth etc. [5]. It is often not easy to separate skin colored objects from non-skin objects like wood, which can appear to be skin colored. Therefore, both skin-color modeling and contours are used to separate the facial region undergoing tracking [1]. The oval shape of the head is often approximated by an ellipse [1][20]. To cope with varying illumination conditions the color model is accommodated over time using the past color distribution and newly extracted distribution from the ellipse's interior. The kernel density based tracking has recently emerged as robust and accurate method due to its robustness to appearance variations and its low computational complexity [4][7][14].

Many color-based tracking approaches assume controlled lighting. In real scenarios an object undergoing tracking may be shadowed by other objects or even by the object itself. Updating the color model is thus one of the crucial issues in color-based tracking. A technique for color model adaptation was addressed in [15]. A Gaussian mixture model was used to represent the color distribution and the linear extrapolation was utilized to adapt the model parameters via a set of labeled training data from a subimage within the bounding box. A non-parametric method that in histogram adaptation employs only pixels which fall in the skin locus was proposed in work [16]. In work [18] the modeling of the color distribution over time is realized through predictive histogram adaptation. Histograms are dynamically updated using affine transformations, warping and resampling. The pixel-wise skin color segmentation is often not sufficient to provide the pixels for color model adaptation because pixels in the image background may also have colors similar with skin colors and this can then lead to over-segmentation. Another issue which should be taken into account is that nearby pixel from skin-colored background may blend with the true skin regions and this can have an adverse effect on subsequent processing of skin regions. The adaptive skin-color filter [6] performs initial skin candidate detection at the beginning and then more accurate tuning of skin model takes place. The adaptation takes into account the skin-like background colors. The method uses HSV color space in which the H coordinates are additionally shifted by 0.5. A comparative study of four state-of-the-art techniques of skin detection under changing illumination conditions can be found in [17].

The proposed algorithm of face tracking under time-varying illumination begins with separating skin and non-skin colors using a database of skin and non-skin pixels. The statistical shape model is utilized in selection of face candidates, yielding the best face candidate on the basis of shape and color criterions. The facial regions are detected during tracking by looking for pixels that have skin

colors. The presence of such pixels in the input image is detected using the skin and non-skin histograms. A kernel color histogram is used in frame-to-frame appearance matching. The detected skin-colored regions are then refined using homogeneity property which exhibit skin regions. In particular, the connected component analysis is applied to label separate regions. Spatial morphological operations for hole and object size filtering are used afterwards. The statistical shape model provides an effective method for fitting oval head shapes to detected face candidates. The algorithm reviews the image focusing the action around the position where the tracked face was detected in the previous frame. The outcome of this stage is a shape fitted to the face candidate. It provides additional information about expected target location. Using outline determined in such a way a local search is conducted to determine the pose of the face allowing for both color and shape information. The key idea of the proposed approach is improved selection of skin pixels to determine the parameters of models expressing the skin evolution over time. Even when a background region situated close to a face region has skin colored pixels, there always exists a boundary between the true skin region and the background. Our aim is to detect such a boundary using Active Shape Models. A second order Markov model is applied to predict the evolution of colors of such skin pixels, gathered in certain number of the last frames. The predicted skin distribution is quantized using a kernel to construct the histogram.

The Active Shape Models, which were originally proposed by Cootes [8], have been modified in work [13] to incorporate color cues. The cited approach does not apply color segmentation to the images. It is based on the minimization of energy functions in the color components. Therefore it admits of only a small change in illumination between two successive frames.

The following section briefly outlines some topics related to statistical shape models. The details of the shape alignment are given in Section 3. Section 4. describes how the Active Shape Model is used to conduct tracking and to support the skin segmentation in video under time-varying illumination conditions. The model of skin colors and their evolution is described in Section 5. Experimental results are shown in Section 6. We report some conclusions in the last section.

2 Facial shape constraints

The method of segmentation and tracking of facial regions proposed in this paper utilizes the statistical shape models. A shape model is utilized to constrain the configuration of a set of candidate skin pixels. An efficient algorithm allows the facial pixels detections to be tested and verified. The method allows for skin detector failures by predicting the locations of missing skin pixels using the shape model. The non-skin pixels outside the shape are not considered in the skin-color model as well.

During shape guided verification of the facial region a set of candidate skin pixels is inspected using shape constraints in two ways. Firstly, a shape model is fitted to the candidate facial region. Secondly, limits are prescribed on the

position, orientation and scale of a set of candidate skin pixels relative to the position, orientation and scale according to their values from the last frame. The aim is to extract pixels belonging only to the tracked face, using the candidate facial mask and the shape constraints. The facial mask is generated from a skin probability image. The skin probability image is extracted on the basis of skin histogram that is accommodated over time.

There are two broad approaches for representing a two-dimensional shape: region-based and contour-based. The region-based methods encode the place occupied by the object through a mask. The methods belonging to this group are sensitive to noise and they cannot cope with partly obscured objects. In contour-based approach the boundary of the object is modeled as an outline. Therefore, such methods can deal better with partially obscured objects and partial occlusions. A contour-based model can be built by placing landmark markers on distinctive features and at some pixels in between. The contour-based instances are usually normalized to canonical scale, translation and rotation in order to make possible comparison among distinct shapes. A distance between corresponding points from the two normalized shapes can be utilized to express the similarity.

Active Shape Models (ASMs or smart snakes) were originally designed as a method for locating given shapes or outlines within images [8]. An ASM-based procedure starts with the mean shape, approximately aligned to the object, iteratively distorts it and refines the pose to obtain a better fit. It seeks to minimize the distance between model points and the corresponding pixels found in the image. A shape consisting of n points can be considered as one data point in $2n-$dimensional space. A classical statistical method for dealing with redundancy in multivariate data is the principal component analysis (PCA). PCA determines the principal axes of a cloud of n points at locations $\mathbf{x}_i$. The principal axes, explaining the principal variation of the shapes, compose an orthonormal basis $\mathbf{\Phi} = \{\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_n\}$ of the covariance matrix $\mathbf{\Sigma} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$. It can be shown that the variance across the axis corresponding to the i-th eigenvalue λ_i equals the eigenvalue itself. An instance of shape can be generated by deforming the mean shape $\bar{\mathbf{x}}$, using a linear combination of eigenvectors $\mathbf{\Phi}$, weighted by so-called modal deformation parameters $\mathbf{b}$. Thus, the new shape can be expressed in the following manner: $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{\Phi}\mathbf{b}$. By varying the elements of $\mathbf{b}$ we can modify the shape. By applying constraints we ensure that the generated shape is similar to the mean shape from the original training data. The deformation of shapes is limited to a subspace spanned by a few eigenvectors corresponding to the largest eigenvalues. We can achieve a trade-off between the constraints on the shape and the model representation by varying the number of eigenvectors.

3 Shape alignment

Given two 2D shapes, $\mathbf{x}_2$ and $\mathbf{x}_1$ our aim is to determine the parameters of a transformation T, which, when applied to $\mathbf{x}_2$ can best align it with $\mathbf{x}_1$ with one-

to-one point correspondence. During alignment we utilize an alignment metric that is defined as the weighted sum of the squares of the distances between corresponding points on the considered shapes. Thus we seek to choose the parameters t of the transformation T to minimize:

$$E = \sum_{i=1}^{n} (\mathbf{x}_{1i} - T_t(\mathbf{x}_{2i}))^T \mathbf{W}_i (\mathbf{x}_{1i} - T_t(\mathbf{x}_{2i})), \tag{1}$$

where $\mathbf{W}$ is a diagonal matrix of weights $\{w_1, w_2, ..., w_n\}$. Expressing T_t in the following form:

$$T_t \equiv \begin{bmatrix} s\cos(\theta) & -s\sin(\theta) & t_x \\ s\sin(\theta) & s\cos(\theta) & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{2i} \\ y_{2i} \\ 1 \end{bmatrix} \tag{2}$$

and denoting $a_x = s\cos(\theta)$, $a_y = s\sin(\theta)$ we can rewrite (1) in the following form: $E = \sum_{i=1}^{n} w_i((a_x x_{2i} - a_y y_{2i} + t_x - x_{1i})^2 + (a_y x_{2i} + a_x y_{2i} - t_y - y_{2i})^2)$. The error E assumes a minimal value when all the partial derivatives are zero. For example, differentiating the last equation with regard to a_x we obtain: $\sum_{i=1}^{n} w_i(a_x(x_{2i}^2 + y_{2i}^2) + t_x x_{2i} + t_y y_{2i}(x_{1i}x_{2i} + y_{1i}y_{2i})) = 0$. Diferentiating w.r.t. remaining parameters and equating to zero gives:

$$\begin{bmatrix} C_1 \\ C_2 \\ X_1 \\ Y_1 \end{bmatrix} = \begin{bmatrix} D & 0 & X_2 & Y_2 \\ 0 & D & -Y_2 & X_2 \\ X_2 & -Y_2 & W & 0 \\ Y_2 & X_2 & 0 & W \end{bmatrix} \begin{bmatrix} t_x \\ t_y \\ a_x \\ a_y \end{bmatrix}, \tag{3}$$

where $X_i = \sum_{k=1}^{n} w_k x_{ik}$, $Y_i = \sum_{k=1}^{n} w_k y_{ik}$, $C_1 = \sum_{k=1}^{n} w_k(x_{1k}x_{2k} + y_{1k}y_{2k})$, $C_2 = \sum_{k=1}^{n} w_k(y_{1k}x_{2k} + x_{1k}y_{2k})$, $D = \sum_{k=1}^{n} w_k(x_{2k}^2 + y_{2k}^2)$, $W = \sum_{k=1}^{n} w_k$. The parameters t_x, t_y, a_x and a_y constitute a solution which best aligns the shapes. An iterative approach to find the minimum of square distances between corresponding model and image points is as follows [8]:

1. Initialize $\mathbf{b}$ to zero.
2. Generate the model points using $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{\Phi b}$
3. Find the pose parameters using (3)
4. Project image pixels $\mathbf{Y}$ into the model co-ordinates using $\mathbf{y} = T_t^{-1}(\mathbf{Y})$
5. Scale $\mathbf{y}$ as follows $\mathbf{y}' = \mathbf{y}/(\mathbf{y}.\bar{\mathbf{x}})$
6. Update $\mathbf{b}$ in the following manner $\mathbf{b} = \mathbf{\Phi}^T(\mathbf{y}' - \bar{\mathbf{x}})$
7. If not converged, go to step 2.

4 Active Shape Model Based Tracking

Tracking can be perceived as a problem of assigning consistent labels to objects being tracked. This is done through maintaining the observations of objects in order to label these so that all observations of a given object in a sequence of images are given the identical label. During shape aligning our algorithm reviews

the binary image focusing the action around the pose that has been determined in the previous frame. The algorithm requires that the new shape center remains within the face mask centered on the previous location of the target. Such an assumption is utilized in kernel based trackers [4][7]. Limits are prescribed on the position, orientation and scale according to their values in the last frame. The binary image is generated in advance on the basis of the skin histogram that is accommodated over time.

The standard ASM aligns the shape model to outlines in an image using only edges. To obtain a rough pose of the face we first utilize the edges of the mask indicating the face area. The final pose of the shape is determined on the basis of the intensity gradient near the edge of the outline, a matching score of colors from the candidate outline and from the outline determined in the previous iteration and the location of the mask. In work [12] a search for the edges in direction perpendicular to the border has been shown optimal. Therefore, a search for the points along profiles normal to the shape boundary has been implemented.

The shape model has been generated using 10 manually segmented images with frontal faces, each represented by 30 characteristic points. The faces have been normalized with regard to orientation and size in order to obtain a set of points with similar physical correspondence across the training collection. All training faces were manually aligned by eyes position.

The oval shape of the head can be reasonably well approximated by an ellipse. Therefore, in this work the model shapes are normalized by aligning the average shape to a fixed circle of landmark points. Such an approach has the advantage that the model can be scaled to size needed be the application through setting only the size of the circle.

Figure 1. demonstrates the performance of the ASM attempting to match the head model to a given binary mask that has been extracted during tracking. To demonstrate the usefulness of statistical shape models in tracking two artifacts at the left and the right side of face border have been manually added. Despite large deformation of shape outline we can observe how precisely the algorithm can align the model shape to the face mask. The shape on the left represents the initial pose that has been utilized in depicted shape alignment. This exemplifies also how the statistical shape models can support the selection of pixels for color model adaptation and thus the prediction of skin evolution over time. The next section is devoted to description of skin-color based image segmentation under time-varying illumination.

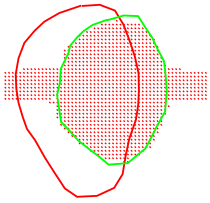

Fig. 1. Shape alignment in presence of artifacts.

5 Skin Color Segmentation under Time-Varying Illumination

The face detection scheme within tracking framework must operate flexibly and reliably regardless of lighting conditions, background clutter in the image, as well as variations in face position, scale, pose and expression. Some tracking applications, for example using a moving camera, do require good detection rates even in case of abrupt changes of illumination. Fast and reliable face segmentation techniques in image sequences are highly desirable capability for many vision systems. Skin color based detection methods are independent to scale, resolution and to some degree of face orientation in the image. A problem with robust detection of skin pixels arises under varying lighting conditions. The same skin patch can look like two different patches under two different conditions. An important issue for any skin-color based tracking system is to provide an accommodation mechanism which could cope with varying illumination conditions that may occur during tracking. In our approach, color distributions are estimated over time and then are predicted under the assumption that lighting conditions vary smoothly over time. The prediction is used to reflect the changing tendency in appearance of the object being tracked. A ground-truth is an evident need during adapting a color model over time to changing illumination conditions [15]. In this approach the evolution of distribution is constrained via statistical shape model and skin locus mechanism. In work [18] the current segmentation and predictions of Markov model were applied to provide a feedback for accommodation. In other work [15] the accommodation process is controlled via mechanism for detecting errors accompanying tracking.

One significant element that should be considered while constructing a statistical model of skin color is the choice of color space. One of the advantages of the HSV color space is that it yields minimum overlap between skin and non-skin distributions. Hue is invariant to certain types of highlights, shadows and shading. A shadow cast does not change significantly the hue color component. It decreases mainly the illumination component and changes the saturation. This color space was utilized in several face detection systems [15][18][19]. The only disadvantage of the HSI color space is the costly conversion from the RGB color space. We handled this problem by using lookup tables.

The histogram is the oldest and most broadly employed non-parametric density estimator. In the standard form it is computed by counting the number of pixels that have given color in region of interest. This operation allows alike colors to be clustered into the separate bin. The quantization into bins reduces the memory and computational requirements. Due to their statistical nature the color histograms can only reflect the content of images in a limited way [21]. Therefore, such representation of color densities is tolerant to noise. Histogram-based techniques are effective only when the number of bins can be kept relatively low and when sufficient data are in disposal [15]. One of the drawbacks of the histogram based density estimation is the lack of convergence to the true density if the data set is small. In certain applications, the color histograms are invariant

to object translations and rotations. They vary slowly under change of angle of view and with change in scale.

The target is represented by the set $S = \{\mathbf{u}_i\}_{i=1}^{N}$, where N is the number of pixels and $\mathbf{u}_i$ denotes vector with HSV components of the $i-$th pixel. Given a set of samples S we can obtain estimate of $p(\mathbf{u})$ using multivariate kernel density estimation [7][9]:

$$p(\mathbf{u}) = p(u^{(1)} = H, u^{(2)} = S, u^{(3)} = V) = \frac{1}{N}\sum_{i=1}^{N}\prod_{l=1}^{3} K_h(u^{(l)} - u_i^{(l)}), \qquad (4)$$

where $K_h(\mathbf{u}) = \frac{1}{(\sqrt{2\pi}h)^d}\exp\left(-\frac{1}{2}\left(\frac{\|\mathbf{u}\|^2}{h^2}\right)\right)$ is a Gaussian kernel of bandwidth h, whereas d denotes the dimension. The quantizatin with 32x32x32 bins has been used to represent both the target as well as the background.

An initial skin histogram, along with the model for non-skin background pixels, has been used in order to compute the probability of every pixel in the first input color image and thus to give the skin likelihood. A model for human skin color distribution was built using a repository of labeled skin pixels that has been prepared in advance. Given the histograms ϕ_{fg} and ϕ_{bg}, the log-likelihood ratio for a pixel with color $\mathbf{u}$ is given by [11]:

$$L(\mathbf{u}) = \max\left(-1, \min\left(1, \log\frac{\max(\phi_{fg}(\mathbf{u}), \delta)}{\max(\phi_{bg}(\mathbf{u}), \delta)}\right)\right), \qquad (5)$$

where δ is a very small number, whereas $\phi_{fg}(\mathbf{u})$, $\phi_{bg}(\mathbf{u})$ denote the frequency of pixels with color $\mathbf{u}$ in the foreground and background, respectively.

Given the probability image the thresholding takes place. After that, the binary image is analyzed using a labeling procedure, which isolates connected components in order to detect the presence of face candidates in the image. Next, the candidate regions are subjected to morphological operations, such as size and hole filtering, to clean up the mask and to generate the mask indicating which pixels belong to the face. After alignment of the model shape with the current mask, the refined face mask is utilized to select from the newly classified pixels the representation of the skin distribution. Using such samples gathered over an initial sequence of frames the sequence-specific motion patterns are learned. A second-order Markov process has been chosen to model the evolution of the color distribution over time [3][18].

Many studies have indicated that the skin tones differ mainly in their intensity value while they form compact cluster in chrominance coordinates [23]. Hence, the evolution of skin cluster can be parameterized at each time instant t by translation, rotation and scaling. The translation parameters $\mathbf{t}_p$ can be extracted on the basis of means from samples constituting a learning distribution, whereas the scaling parameters $\mathbf{s}_p$ can be estimated from their standard deviations. The eigenvectors of the covariance matrices of samples from two consecutive frames define two coordinate frames, which can be then used to estimate the rotations $\mathbf{r}_p$ [18].

The work [3] demonstrated that affine motion can be described via a second-order auto-regressive Markov process:

$$\mathbf{X}(t_{k+1}) - \bar{\mathbf{X}} = A_2(\mathbf{X}(t_{k-1}) - \bar{\mathbf{X}}) + A_1(\mathbf{X}(_{t_k}) - \bar{\mathbf{X}}) + B_0\mathbf{w}_k, \tag{6}$$

where $\mathbf{X} = \{\mathbf{t}_p^T,\ \mathbf{s}_p^T,\ \mathbf{r}_p^T\}$ is the vector parameterizing the skin evolution. The parameters which should be learned are A_0, A_1, and $C = BB^T$ because B cannot be observed directly. It was shown in [2] that the matrices A_0 and A_1 can be estimated on the basis of the following equations:

$$S_{20} - \hat{A}_0 S_{00} - \hat{A}_1 S_{10} = 0 \tag{7a}$$
$$S_{21} - \hat{A}_0 S_{01} - \hat{A}_1 S_{11} = 0, \tag{7b}$$

where $S_{ij} = \sum_{k=1}^{m-2} \left(\mathbf{X}(t_{(k-1)+i})\mathbf{X}^T(t_{(k-1)+j})\right),\quad i,j = 0,1,2$, and m denotes number of learning frames. Having A_0 and A_1 we can estimate C from the following equation: $\hat{C} = \frac{1}{m-2}Z(A_0, A_1)$, where $Z(A_0, A_1) = S_{22} + A_1 S_{11} A_1^T + A_0 S_{00} A_0^T - S_{21} A_1 - S_{20} A_0^T + A_1 S_{10} A_0^T - A_1 S_{12} - A_0 S_{02} + A_0 S_{01} A_1^T$.

On the basis of predicted distribution the histogram $\phi_{fg^{(p)}}$ of skin colors is extracted. After normalization of the histogram we perform an adaptation which combines the histogram that had been obtained from the predicted distribution and the histogram from the last frame. Adaptation is made according to the following equation:

$$\phi_{fg^{(u)}}(t) = (1-\alpha)\phi_{fg}(t-1) + \alpha\phi_{fg^{(p)}}(t), \tag{8}$$

where the adaptation coefficient α has been determined empirically. The histogram $\phi_{fg^{(u)}}(t)$ has been subjected to segmentation procedure to produce the face mask. The refined face mask by statistical shape model, as discussed in Section 4., has been then used to collect the newly classified skin pixels in a list.

The refined face mask by statistical shape model can contain non-skin pixels. Experiments demonstrated that the part of face below the hair was a source of such inadequate pixels. To deal with this undesirable effect, the pixels collected in the mentioned above list were additionally inspected if they fall within the prepared in advance skin locus. A prepared off-line two-dimensional table defining possible skin chromaticities has been used at this stage. It has shown to be useful especially in eliminating non-skin pixels from the representation of the skin distribution in a sudden change of illumination.

The list prepared in such a way has been utilized to generate the histogram $\phi_{fg^{(n)}}$. Finally, this histogram has been updated in the following manner:

$$\phi_{fg}(t) = (1-\alpha)\phi_{fg}(t-1) + \alpha\phi_{fg^{(n)}}(t). \tag{9}$$

This histogram has been utilized to generate the skin image probability during tracking.

6 Experiments

To test the proposed method of face tracking we performed various experiments on real images. We utilized the Carphone sequence as our first test set. The model of the face shape was prepared on images not containing the face from the considered test sequence. Through this sequence we want to highlight the behavior of the tracking algorithm in case of errors in target segmentation. We can notice in Fig. 2. that even if the segmentation does not separate the object of interest from the background, the contour generated from the active shape model supports greatly the extraction of the object. For example, the images from the second column demonstrate that without the ASM based shape refinement the color model would be influenced by the hair colors as well as by the bow-tie colors, see also frame #176.

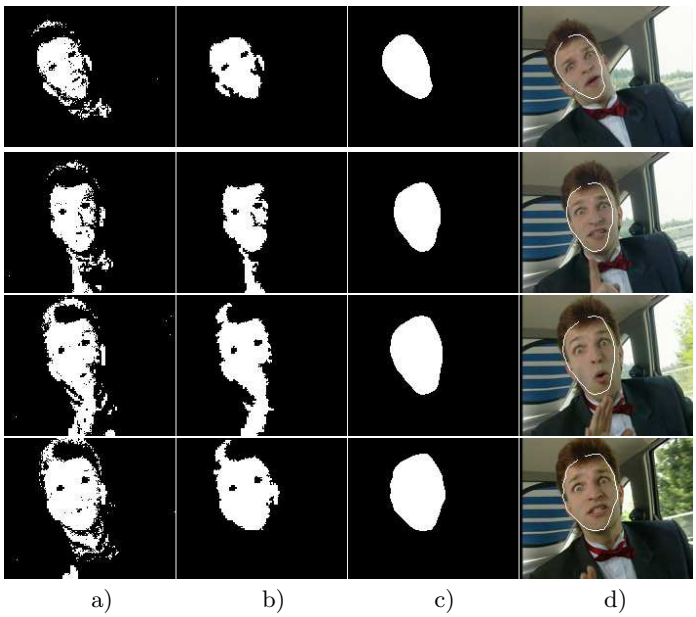

Fig. 2. Face tracking using the Carphone test sequence, frames #84, #125, #176, #200 (from top to bottom). Binary image (a). Hole and object size filtering (b). ASM-based refinement of the face mask (c). Input image with the extracted face outline (d).

To study the adaptation performance in time-varying illumination conditions we conducted experiments with two configurations of the tracking algorithm. In the first configuration we utilized the predictions of the skin evolution, whereas in the second one only the newly classified pixels have been used to accommodate the histogram. The number of learning images has been set to 20. Typically in almost 90% images, the predictions lead to better fidelity in approximation of the face, see also Fig. 3. In particular, the first configuration detected smaller number of skin-pixels in the background in all images, compare Fig. 3a with

Fig. 3c. Other tracking results using this image sequence can be found in [1][5]. The presented system runs at 176x144 image resolution at frame rates of 12-15 Hz on a 2.4 GHz PC. The algorithm has also been tested with Claire and Foreman test sequences as well as PETS-03 meeting recordings. The superior tracking performance over face tracking algorithm using intensity gradients and kernel histograms was observed in all above mentioned sequences. Particularly, smaller "jumps" of the shape indicating the face location from frame to frame have been perceived. In varying illumination can arise the superiority over Mean-Shift because of reduced adaptation capabilities of Mean-Shift methods.

a) b) c) d)

Fig. 3. Tracking in time-varying illumination, frames #301, #310, #319, #330 (from top to bottom). Histogram update using: predictions (a,b), newly classified pixels (c,d).

7 Conclusion

A face detection and tracking strategy has been described. The accommodation of the skin histograms over time takes place using feedback from shape, newly classified skin pixels and predictions of the skin color evolution. Once a face is being tracked, the color model adapts according to changes in appearance and therefore improves tracking performance.

Acknowledgment

This work has been supported by MNSzW within the project 3 T11C 057 30.

References

1. Birchfield, S.: Elliptical Head Tracking Using Intensity Gradients and Color Histograms, In Proc. IEEE Conf. on Comp. Vis. and Patt. Rec., (1998) 232–237

2. Blake, A., Isard, M., Reynard, D.: Learning to Track the Visual Motion of Contours, Artificial Intelligence, **78** (1995) 101–133
3. Blake, A., Isard, M.: Active Contours, Springer (1998)
4. Bradski, G. R.: Computer Vision Face Tracking as a Component of a Perceptual User Interface, In Proc. IEEE Workshop on Appl. of Comp. Vision (1998) 214–219
5. Chen, Y., Rui, Y., Huang, T.: Mode-based Multi-Hypothesis Head Tracking Using Parametric Contours, In Proc. IEEE Int. Conf. on Aut. Face and Gesture Rec. (2002) 112–117
6. Cho, K. M., Jang, J. H., Hong, K. S.: Adaptive Skin Color Filter, Pattern Recognition, **34**(5) (2001) 1067-1073
7. Comaniciu, D., Ramesh, V., Meer, P.: Real-Time Tracking of Non-Rigid Objects Using Mean Shift, In Proc. IEEE Conf. on Comp. Vis. Patt. Rec. (2000) 142–149
8. Cootes, T.: An Introduction to Active Shape Models, Model-Based Methods in Analysis of Biomedical Images, [in:] Image Processing and Analysis, Eds., R. Baldock and J. Graham, Oxford University Press (2000)
9. Elgammal, A., Duraiswami, R., Davis L. S.: Probabilistic Tracing in Joint Feature-Spatial Spaces, In Proc. IEEE Conf. on Comp. Vis. and Patt. Rec. (2003) 16–22
10. Fieguth, P., Terzopoulos, D.: Color-Based Tracking of Heads and Other Mobile Objects at Video Frame Rates, In Proc. IEEE Conf. on Comp. Vis. and Patt. Rec., Hilton Head Island (1997) 21–27
11. Han, B., Davis, L.: Robust Observations for Object Tracking, In Proc. Int. Conf. on Image Processing (2005) 442–445
12. Isard, M., Blake, A.: Contour Tracking by Stochastic Propagation of Conditional Density, European Conf. on Computer Vision, Cambridge (1996) 343–356
13. Koschan, A., Kang, A., Paik, J., Abidi, B., Abidi, M.: Color Active Shape Models for Tracking Non-Rigid Objects, Pattern Recognition Letters, **24** (2003) 1751–1765
14. Perez, P., Hue, C., Vermaak, J., Gangnet, M.: Color-Based Probabilistic Tracking, European Conf. on Computer Vision (2002) 661–675
15. Raja, Y., McKenna, S. J., Gong, S.: Color Model Selection and Adaptation in Dynamic Scenes, Proc. European Conf. on Computer Vision (1998) 460–474
16. Soriano, M., Martinkauppi, B., Huovinen, S., Laaksonen, M.: Adaptive Skin Color Modelling Using the Skin Locus for Selecting Training Pixels, Pattern Recognition, **36** (2003) 681–690
17. Soriano, M., Martinkauppi, B., Pietikainen M., Detection of Skin under Changing Illumination: A Comparative Study, Int. Conf. on Image Analysis and Proc. (2003) 652–657
18. Sigal, L., Sclaroff, S., Athitsos, V.: Estimation and Prediction of Evolving Color Distributions for Skin Segmentation under Varying Illumination, In Proc. IEEE Conf. on Comp. Vis. and Patt. Rec. (2000) 2152–2159
19. Sobottka, K., Pitas, I.: Segmentation and Tracking of Faces in Color Images, In Proc. of the Sec. Int. Conf. on Aut. Face and Gesture Rec. (1996) 236–241
20. Srisuk, S., Kurutach, W., Limpitikeat, K.: A Novel Approach for Robust Fast and Accurate Face Detection, Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems, **9**(6) (2001) 769–779
21. Swain, M. J., Ballard, D. H.: Color Indexing, Int. J. of Comp. Vision **7**(1) (1991) 11–32
22. Yang, M.-H., Krigman, D., Ahuja, N.: Detecting Faces in Images: A survey, IEEE Trans. on Pattern Analysis and Machine Intelligence, **24**(1) (2002) 34–58
23. Yang, J., Weier, L., Waibel, A.: Skin-Color Modelling in Color Images, In Proc. Asian Conf. on Computer Vision (1998) II:687–694