We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Tracking of Facial Regions Using Active Shape Models and Adaptive Skin Color Modeling

Bogdan Kwolek
Rzeszow University of Technology
Poland

1. Introduction

It is widely accepted that skin-color is an effective and robust cue for face detection, localization and visual tracking. Well-known methods of color modeling, such as histograms and Gaussian mixture models enable creation of appropriately exact and fast detectors of skin. In particular, skin color-based methods are robust to changes in scale, resolution and partial occlusion. In real scenarios an object undergoing tracking may be shadowed by other objects or even by the object itself. However, many color-based tracking approaches assume controlled lighting. These methods construct or learn models in advance and then use them in tracking, without adaptation to suit new conditions. Consequently, these techniques usually fail or have significant drifts after some period of time, mainly due to variation of lighting in the surrounding. Thus, 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 conditions vary. 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. Two types of information are typically used to perform segmentation during face tracking. The first is color information (Bradski, 1998; Comaniciu et al., 2000; Fieguth & Terzopoulos, 1997; Perez et al., 2002; Sobottka & Pitas, 1996). The second is the geometric configuration of the face shape (Chen et al., 2002). 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 (Birchfield, 1998). The oval shape of the head is often approximated by an ellipse (Birchfield, 1998; Srisuk et al., 2001). 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. However, such tracker pays little attention to what lies inside the ellipse and what is utilized to accommodate the color model. 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 (Bradski, 1998; Comaniciu et al., 2000; Perez et al., 2002). Due to the use of a simple pixel-based representation as well as reduced adaptation capabilities of Mean-Shift methods the algorithm performs poorly under large illumination change.

Updating the color model is one of the crucial issues in color-based tracking. A technique for color model adaptation was addressed in (Raja et al. 1998). A Gaussian mixture model was

Source: Scene Reconstruction, Pose Estimation and Tracking, Book edited by: Rustam Stolkin,
ISBN 978-3-902613-06-6, pp.530, I-Tech, Vienna, Austria, June 2007

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 (Soriano et al., 2003a). In work (Sigal et al., 2000) 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 select the pixels for adaptation of a color model 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 (Cho et al., 2001) performs initial skin candidate detection at the beginning and then more accurate tuning of a skin model takes place. The adaptation takes into account the skin-like background colors. The method uses the 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 (Soriano et al., 2003b).

A few attempts have been proposed to track objects under large change in illumination (Hager & Belhumeur, 1996; La Cascia et al., 2000). These algorithms follow the same idea consisting in the usage of a low dimensional linear subspace to approximate the space of all feasible views of the object under different lighting conditions. To perform the tracking one needs to construct the basis images from a set of images collected at fixed pose under different lighting conditions.

The *key* idea of the proposed approach is an improved selection of pixels to determine the parameters of models expressing the evolution of skin color 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 delineate such a boundary under varying illumination conditions by means of Active Shape Models. In context of dealing with skin-color segmentation under time varying illumination the Gabor filters are particularly useful as they are robust to variability in images arising due to variation in lighting and contrast. Active Shape Models (ASM) were originally proposed by Cootes (Cootes, 2000). They allow for considerable variability of instances of models represented in a subspace spanned by eigenvectors.

The algorithm for segmenting and tracking a face in a sequence of color images enables reliable segmentation of facial region during face tracking 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. The detected skin-colored regions are then refined with regard to spatio-temporal coherence. The algorithm reviews the image focusing the action around the location of the face in the previous frame. In particular, the connected component analysis is applied in the binary image to label separate regions. Spatial morphological operations for hole and object size filtering are used afterwards. Using prior knowledge about the target shape the Active Shape Model seeks to match a set of model points to the image. While interpreting the image contents we employ statistical shape models built on intensity gradients, distance between color of pixels in subsequent frames and the phase of Gabor filter responses. In the

first iteration we always utilize the distance to the edge of extracted in advance facial mask to find a plausible starting configuration. The coherence score between corresponding characteristic points, which is determined using phase of the Gabor filter responses, improves considerably the tracking capabilities of the method. The outcome is a shape fitted to the tracked face.

The user only needs to initialize the tracker in the first frame. After a fixed number of frames the tracker automatically switches from tracking with the learning phase to the model-based tracking. A second order Markov model is applied to predict the evolution of colors of skin pixels, gathered within shape interiors in certain number of the last frames. During the tracking, the matching are not performed between only image pairs, but also between the current frame and the shape model. The accommodation of the skin histogram over time takes place on the basis of feedback from shape, newly classified skin pixels and predictions of the skin color evolution.

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 in our system to conduct tracking and to support the skin segmentation. It presents in detail all ingredients of our ASM-based tracker and reports results, which were obtained in experiments with various cues. The model of skin colors and their evolution is described in Section 5. Experiments conducted in varying illumination are described as well.

2. Point Distribution Model

The method for segmentation and tracking of facial regions, which is presented in this chapter 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 detection of facial pixels to be tested and verified. Thus, it deals with failures of a skin detector. The non-skin pixels that are placed outside of the shape are not considered in the skin-color model.

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, intensity gradient, coherence of the phase of Gabor filter responses, and the shape constraints. The facial mask is generated from a skin probability image. The skin probability image is extracted on the basis of a 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 better deal 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 between them.

Active Shape Models (ASMs or smart snakes) were originally designed as a method for locating given shapes or outlines within images (Cootes, 2003). An ASM-based procedure starts with the base shape, approximately aligned to the object, iteratively distorts it and refines its 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 npoints 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. By deforming the mean shape $\bar{\mathbf{x}}$, using a linear combination of eigenvectors $\mathbf{\Phi}$, weighted by so-called modal deformation parameters **b**, we can generate an instance of the shape. Therefore, the new shape can be expressed in the following manner: $\mathbf{x}=\bar{\mathbf{x}}+\mathbf{\Phi b}$. By varying the elements of 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. Through applying limits of $\pm 3\sqrt{\lambda_i}$ to each element b_i of **b**, where λ_i is the variance of the *i* - th parameter b_i , we can operate on plausible values of **b**. The deformation of the shape is constrained 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. If all principal components are employed, ASM can represent any shape and no prior knowledge about the shape is utilized.

3. Shape Alignment

Given two 2D shapes, $\mathbf{x}_1$ and $\mathbf{x}_2$ 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:

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

where **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(θ), a_y =s sin(θ) we can rewrite (1) in the following form:
$E=\sum_{i=1}^{n}w_i\left((a_x x_{2i}-a_y y_{2i}+t_x-x_{1i})^2+(a_y x_{2i}+a_x y_{2i}-t_y-y_{2i})^2\right)$. The error *E* assumes a minimal value when all the partial derivatives are zero. Differentiating the last equation with regard

to a_x we obtain: $\sum_{i=1}^{n} w_i\left(a_x(x_{2i}^2+y_{2i}^2)+t_x x_{2i}+t_y y_{2i}(x_{1i}x_{2i}+y_{1i}y_{2i})\right)=0$. 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 $X_k=\sum_{i=1}^{n} w_i x_{ki}$, $Y_k=\sum_{i=1}^{n} w_i y_{ki}$, $C_1=\sum_{i=1}^{n} w_i(x_{1i}x_{2i}+y_{1i}y_{2i})$, $C_2=\sum_{i=1}^{n} w_i(y_{1i}x_{2i}+x_{1i}y_{2i})$, $D=\sum_{i=1}^{n} w_i(x_{2i}^2+y_{2i}^2)$, $W=\sum_{i=1}^{n} w_i$. 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 (Cootes, 2003):

1. Initialize shape parameter **b** to zero.
2. Generate the model instance $\mathbf{x}=\bar{\mathbf{x}}+\mathbf{\Phi b}$.
3. Find the pose parameters using (3), which best map **x** to **Y**.
4. Invert the pose parameters and then use to project image pixels **Y** into the model co-ordinate frame: $\mathbf{y}=T_t^{-1}(\mathbf{Y})$.
5. Project **y** into the tangent plane to $\bar{\mathbf{x}}$ through scaling it by $1/(\mathbf{y}.\bar{\mathbf{x}})$: $\mathbf{y}'=\mathbf{y}/(\mathbf{y}.\bar{\mathbf{x}})$.
6. Update **b** to match **y**′ as follows: $\mathbf{b}=\mathbf{\Phi}^{\mathbf{T}}(\mathbf{y}'-\bar{\mathbf{x}})$.
7. If not converged, repeat starting from 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 there is an overlap between the image region occupied by the object in the previous iteration and the new object region. Such an assumption is utilized in Mean-Shift trackers (Bradski, 1998; Comaniciu et al. 2000), which require significant overlap on the target kernels in consequent frames. In our system limits are prescribed on the position, orientation and scale of the target according to their values in the last frame. The binary image is generated prior to shape fitting 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 contours. It works well on images with consistent shape and appearance. It requires good initialization and is inadequate when the shape variations are highly non-linear. To cope with such constraints we initialize the locating of the face in each frame by the use of the binary mask. Its boundary indicates a rough location as well as shape of the face. In work (Koschan et al., 2003) an incorporation of color cues into the ASM framework has been proposed. However, the mentioned approach does not apply color segmentation. 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.

Fig. 1. demonstrates the performance of the ASM attempting to match the head model to a given binary mask that has been extracted on the basis of color model. 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 the shape outline we can observe how precisely the algorithm can align the shape to such a face mask. The shape on the left is the base shape in the initial pose that has been utilized in depicted shape alignment. This figure 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.

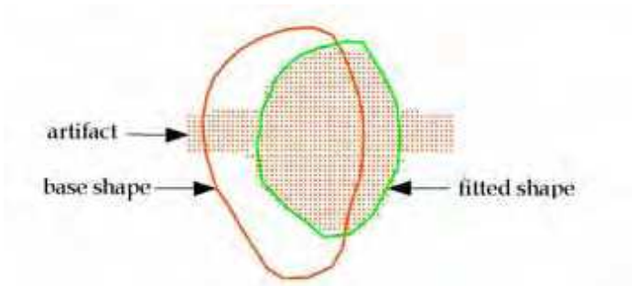


Fig. 1. Shape alignment in presence of manually added artifacts to extracted facial mask.

The shape model has been prepared on the basis of 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 eye position.

The oval shape of the head can be reasonably well approximated by an ellipse. During preparing the statistical model of the head shape 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 a needed size via setting only the size of the circle. The pose of the shape during the tracking is determined on the basis of the distance to the edge of face mask, intensity gradient near the edge of the outline, matching score of colors from the candidate outline and from the outline determined in the previous iteration, and phase of the Gabor filter responses. In the following subsections we present how each of the mentioned above cues contributes to the cost function that is calculated during searching for the best fit to the tracked face.

4.1. Distance to the Edge of the Facial Mask

In work (Isard & Blake, 1996) a search for the edges in direction perpendicular to the shape border has been shown as optimal. Therefore, a search for the points along profiles normal to the shape border is employed in our system. Fig. 2. demonstrates sample shape and location of the normals corresponding to characteristic points of our face representation.

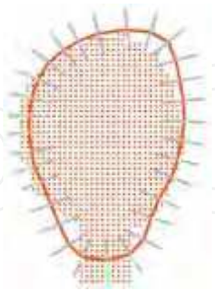


Fig. 2. The location of the normals to base shape.

Fig. 3. shows some shapes that were determined using the distance to the edge of the binary mask undergoing fitting. The binary images indicating skin color like areas were extracted on the basis of histograms accommodated over time. The experiments were conducted in home/office environment in front of wooden doors and a piece of furniture.

Fig. 3. ASM-based face tracking using distance to the edge of the face mask. Frames #1, #2, #20, #40, #60, #80, #100 and #120 (from left to right and from top to bottom). Left images in the pairs are binary ones, whereas right images depict the outlines fitted to the face.

The candidates of facial region are extracted on the basis of histograms modeling the distribution of skin color. Histograms are accommodated over time from newly classified skin pixels and predictions of the skin-color evolution. The backprojected histograms are employed to generate binary images. Such images are then used in determining the connected components. Spatial morphological operations, such as size and hole filtering are employed next. Using the location of the face in the previous frame, a single binary component is extracted finally. The Active Shape Model seeks to match a set of model points to such a facial mask. In the images shown above we can perceive that only the usage of distance to the edge of the mask can lead to shapes that are well fitted to the face. The results demonstrate that the mask can be very useful in the initialization of the shape fitting.

4.2. Intensity Gradient

Figure 4 demonstrates some results that were achieved using intensity gradient while shape fitting to the tracked face. We apply the binary mask in the first iteration that initializes the matching of the set of model points to the edges. The gradient magnitude is calculated on the basis of the Sobel mask. The filtering with Gaussian mask precedes the extraction of the intensity gradient. The search is done along lines perpendicular to the shape.

Fig. 4. ASM-based face tracking, using gradient. Frames #1, #2, #20, #60, #80, #100, #120.

Comparing the extracted shapes at Fig. 3. and Fig. 4. we can observe that the shapes generated on the basis of intensity gradient tend to fit the upper part of the person's head. Such an effect occurs because the strongest edge is not always the object edge. The number of pixels indicating skin like areas in the background is slightly larger in images shown in Fig. 4.

4.3. Intensity Gradient and Color

As demonstrated in (Birchfield, 1998), the contour cues combined with color can be very useful to distinguish the tracked head when both a model of the color distribution and the elliptical model are accommodated over time. When the contour information is poor or is temporary unavailable, color information can be very useful alternative to extract the tracked object. Some tracking results that were obtained using color and intensity gradient cues are depicted in Fig. 5. As in previous experiments, the searching starts from the final location in the previous frame and proceeds iteratively to find the best fit of the shape to the face. In the first iteration the distance to the edge of the face mask is employed. The incorporation of information about the temporal coherence of color results in tracking with small shape's jumps, even in the presence of skin like colors in the background.

Fig. 5. ASM-based face tracking using intensity gradient and color. Frames #1, #2, #20, #40, #60, #80, #100 and #120.

4.4. Phase of Gabor Filter Responses

In order to improve further the quality of shape fitting we exploit Gabor filter responses. This choice is biologically motivated since it has been shown that they model the response the human cortical cells, which are both orientation and frequency selective. In context of dealing with skin-color segmentation under time varying illumination, the Gabor filters can be particularly useful as they are robust to variability in images arising due to variation in lighting and contrast.

A 2-D Gabor filter is created by modulating a 2-D sine wave with a Gaussian envelope. The 2-D kernel of the Gabor filter is given by:

$$g(x,y,\theta_k,\lambda)=\exp\left[-\frac{(x\cos\theta_k+y\cos\theta_k)^2}{2\sigma_x^2}-\frac{(-x\sin\theta_k+y\cos\theta_k)^2}{2\sigma_y^2}\right]\exp\left\{\frac{2\pi(x\cos\theta_k+y\sin\theta_k)}{\lambda}\right\} \quad (4)$$

where σ_x and σ_y denote the standard deviations of the Gaussian envelope along the x and y, respectively, whereas λ and θ are the wavelength and orientation of the 2-D sine wave, respectively. The spread of the envelope is determined via the sine wavelength λ . θ_k is defined as follows: $\theta_k=\frac{\pi}{n}(k-1)$, where k =1, 2,..., n and n represents the number of the considered orientations.The Gabor filter response is calculated by convolving the filter kernel specified by θ_k and λ with the gray-level image I :

$$O(x,y,\theta_k,\lambda)=I(x,y)*g(x,y,\theta_k,\lambda)\ . \quad (5)$$

Fig. 6. shows the real part of Gabor filtered images. The images show the advantages of multiscale image representation-based on Gabor functions in feature matching. In results shown here, we have used four scales and four orientations in representing the landmark points. In our system we employ the efficient Gabor filter implementation of Nestares (Nestares et al., 1998). This pyramidal multiscale Gabor transform that allows very efficient implementation in the spatial domain is faster than conventional FFT implementations.

Given a characteristic point of our shape model we are interested in a correspondence score between the considered pixel at the normal and the corresponding pixel that has been acquired at the initial outline. Such a correspondence score can be estimated using the phase of the Gabor filter responses. Suppose that for a Gabor filter with orientation θ and wavelength λ the phase at a point x_t is $\phi_{\lambda,\theta}$. Given a response of a single filter the similarity between points x_t and x_1 is proportional to $\exp(-(\phi_{\lambda,\theta}(\mathbf{x}_t)-\phi_{\lambda,\theta}(\mathbf{x}_1))^2)$. The matching score between points x_t and x_1 can be computed in the following manner:

$$G(\mathbf{x}_1,\mathbf{x}_t)=C_h\prod_{\lambda,\theta}\left(\exp(-(\phi_{\lambda,\theta}(\mathbf{x}_t)-\phi_{\lambda,\theta}(\mathbf{x}_1))^2)+1\right), \quad (6)$$

where C_h is a normalization constant ensuring that G varies between 0 and 1. By adding 1 to each factor during multiplication we limit the predominance of a single filter in the filter outcome.

Fig. 7. illustrates the coherence score between the landmark points that were acquired from the shape in the first frame and pixels from the frame #10. For visualization purposes a face subimage is included in the probability image. The brighter the pixel representing probability is, the higher is the coherence probability. The images demonstrate the

usefulness of phase in precise alignment the shape to the facial landmarks. The location of the landmark points for which the coherence probability has been computed can be found at Fig. 8c.

To achieve a better fit of the model shape to image data the method elaborated by (Cootes, 2000) uses searching profiles. Within such profiles this method looks for a sub-profile with statistics that best match the training profile. A representation of the training profile of each landmark is constructed off-line using a collection of the gray level values along the search profiles. The best match is determined by searching for a sub-profile for which a square error function takes the minimal value. The searching starts at the top level of the multi-resolution pyramid and continues at the lover level using the search outcome of the previous level. However, this method is sensitive to changes in illumination. One of the main advantages of our method is its robustness to variations in illumination and contrast. Our method does not require an off-line training stage and takes also the advantages of multiresolution analysis.

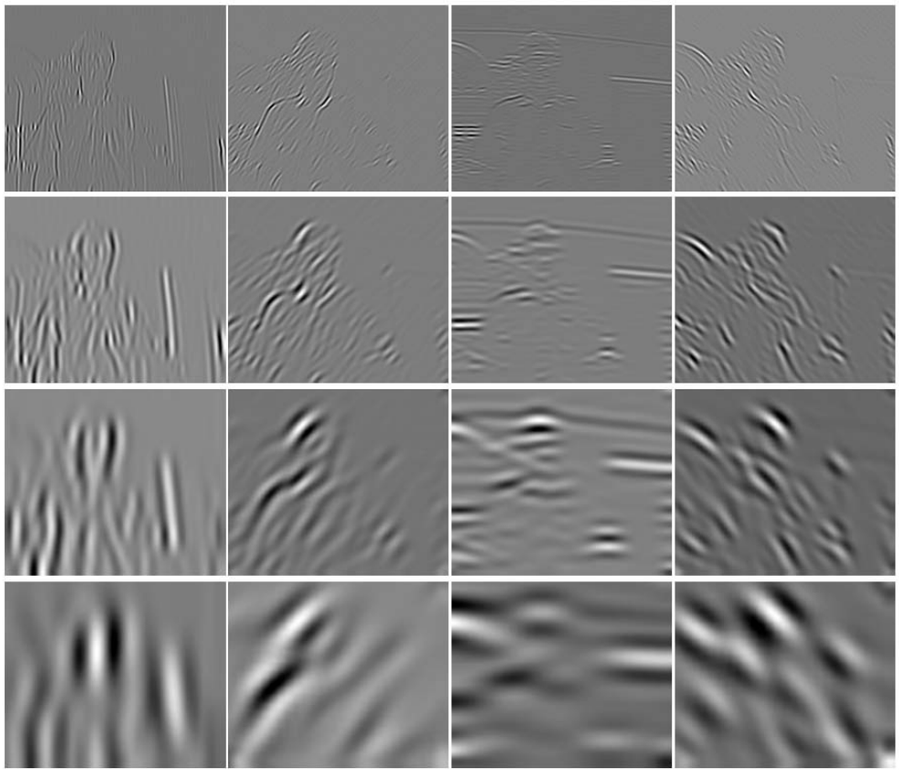


Fig. 6. Gabor decomposition of the test image.

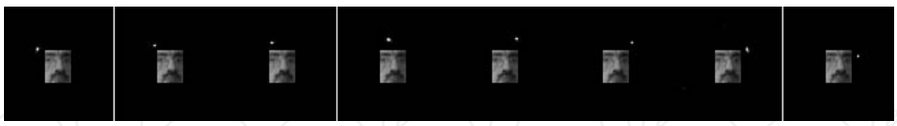


Fig. 7. Gabor filter-based coherence score between the pixels located at landmark points of the shape fitted to face in frame #1 and image pixels from frame #10.

The initialization of the tracker begins with a separation of skin and non-skin colors, see Fig. 8a, using a database of skin and non-skin pixels. The face mask obtained in such a way is utilized to determine the initial pose of the base shape, see Fig. 8b. Experiments demonstrated that such a rough initialization is sufficient to conduct successful tracking in typical scenarios. Good choices for reference pixels to compute the phase score are points at corners or borderlines. Pixels located at the borderline between the shirt and the face are examples of such pixels too. Therefore, after the automatic determination of the shape pose we manually correct the pose of the shape in order to place some of the landmark points of the shape at mentioned above points. Fig. 8c illustrates a typical fit of the base shape to the face after manual correction of the pose. It has been obtained through clock-wise rotation of the shape depicted in Fig. 8b. Although our algorithm does not require very precise initialization, a far more precise initial fit of the shape to the face can be obtained. In case of such a need our graphical interface provides sufficient support and flexibility. For example, we can choose a mode of variation and its weight and then visualize the generated shape. In particular, thanks to such functionality we can determine the number of eigenvalues that are needed to approximate any tracking example within a given accuracy. After specifying the max weight and step we can animate the deformations of the shape in front of the face. This helps in selecting a set of parameters preventing the algorithm from convergence to an unrealistic shape. In another option of the program, through a specification of the weight for each mode we can observe deformation of the shape and its fit to the face. The mentioned above functionality acknowledged also its usefulness at the training stage.

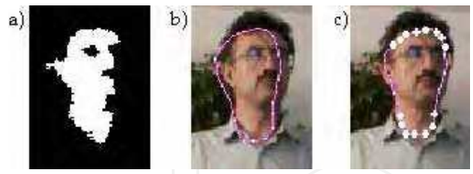


Fig. 8. Initialisation of the tracker.

Fig. 9. illustrates some tracking results that were obtained using the distance to the edge of mask, the intensity gradient and the phase coherence. Through this sequence we want to highlight an improved fit of the shape to the face while the tracking. Comparing images from this figure with corresponding images from Fig. 3.-5. we can notice that thanks to Gabor filter responses, the upper part of the shape is located almost in all frames at the border between the face and hairs of the person's head. A similar effect consisting in a close location of the bottom part of the outline to the face-shirt boundary can also be observed. The accuracy of locating the boundary of the face is constrained by the assumed shape model.

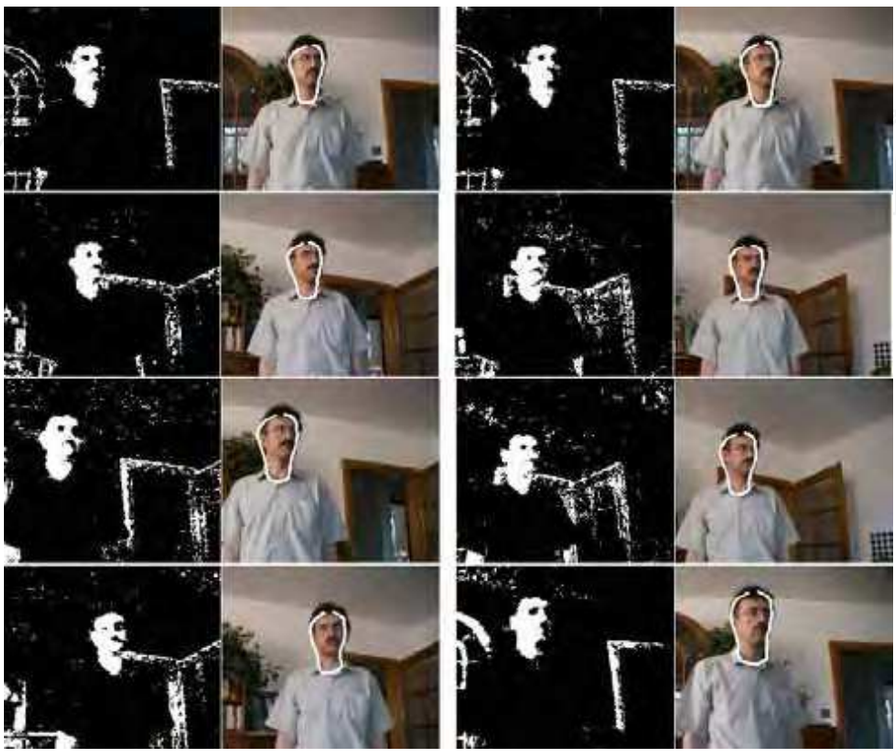


Fig. 9. ASM-based face tracking using the distance to the edge of face mask, intensity gradient and coherence of the phase. Frames #1, #2, #20, #40, #60, #80, #100 and #120.

In the experiments described above we used two modes to approximate the oval shape of the human head. We constructed also shape models with increasing number of modes in order to test their ability to approximate the outline of the face as well as their ability to generalize. The shapes we can obtain arise via linear combinations of the shapes seen in a training set. Thus, the examples of the training repository have also an influence on the approximation as well as generalization capabilities. Some results from the tracking experiments using four modes are presented in Fig. 10. An improved fit to the face can be observed. Our experimental findings show that the 2-3 nodes provide sufficient approximation having on regard comparable face sizes in the image. The model employed in this work has been utilized in our former work (Kwolek, 2006). It has been prepared on the basis of images not containing the faces from the presented here test sequences. A very simple model built on landmark points constituting a shape like an egg can be sufficient to approximate the oval shape of the head in many tracking scenarios. The number of landmark points can be smaller as well. The model parameterized by the number of landmarks, which we decided to use in our experiments provides sufficient approximation for faces occupying larger areas of image.

Fig. 10. ASM-based face tracking using distance to the edge of the face mask, intensity gradient and coherence of the phase. The number of modes is set to four. Each 10-th frame of 120 frames long sequence is presented.

The presented above experimental results were achieved using 10 iterations and they were conducted in front of wooden doors and a piece of furniture. Large shape deformations are made in the first few iterations, which give the scale and shape roughly correct, see Fig. 11b-d. While the searching progresses the deformations are smaller. Fig. 11c demonstrates the shape in first iteration, whereas Fig. 11d shows shapes in 9-th and 10-th iteration, respectively. The images depict also how the statistical shape models can support the selection of pixels for adaptation of the skin model. The skin-color based image segmentation under time-varying illumination is described in next section.

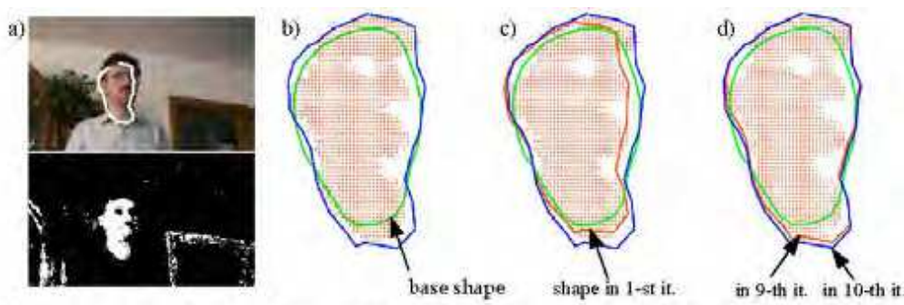


Fig. 11. Examples of search. Input and binary images (a). Process of searching (b-d).

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 (Raja et al., 1998). In this approach the evolution of distribution is constrained via statistical shape model and skin locus mechanism. In work (Sigal et al., 2000) the current segmentation and predictions of Markov model were applied to provide a feedback for accommodation. In other work (Raja et al., 1998) 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 (Raja et al., 1998; Sigal et al., 2000; Sobottka & Pitas, 1996). 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 (Swain & Ballard, 1991). 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 (Raja et al., 1998). 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 (Comaniciu et al., 2000; Elgammal et al., 2003):

$$p(\mathbf{u}) = p\left(u^{(1)} = H, u^{(2)} = S, u^{(3)} = V\right) = \frac{1}{N}\sum_{i=1}^{N}\prod_{l=1}^{3} K_h\left(u^{(l)} - u_i^{(l)}\right), \tag{7}$$

where $K_h(\mathbf{u}) = \frac{1}{(\sqrt{2\pi}h)^d}\exp\left(-\frac{\|\mathbf{u}\|^2}{2h^2}\right)$ is a Gaussian kernel of bandwidth h, whereas d denotes the dimension. The quantization 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 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 (Han & Davis, 2005):

$$L(\mathbf{u}) = \max\left(-1, \min\left(1, \log\frac{\max(\phi_{fg}(\mathbf{u}), \delta)}{\max(\phi_{bg}(\mathbf{u}), \delta)}\right)\right), \tag{8}$$

where δ is a very small number, whereas $\phi_{fg}(\mathbf{u})$ and $\phi_{bg}(\mathbf{u})$ 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 via 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 (Blake & Isard, 1998; Sigal et al., 2000).

Many studies have indicated that the skin tones differ mainly in their intensity value while they form compact cluster in chrominance coordinates (Yang et al., 1998). 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 $\mathbf{r}_p$.

The work (Blake & Isard, 1998) demonstrated that affine motion can be described via a second-order auto-regressive Markov process:

$$\mathbf{X}(t+1) - \overline{\mathbf{X}} = A_2\left(\mathbf{X}(t_{k-1}) - \overline{\mathbf{X}}\right) + A_1\left(\mathbf{X}(t_k) - \overline{\mathbf{X}}\right) + B_0\mathbf{w}_k, \tag{9}$$

where $\mathbf{X} = \{\mathbf{t}_p^{\mathrm{T}}, \mathbf{s}_p^{\mathrm{T}}, \boldsymbol{t}_p^{\mathrm{T}}\}$ is the vector parameterizing the skin evolution. The parameters which should be learned are $\mathbf{A}_0$, $\mathbf{A}_1$ and $C = BB^{\mathrm{T}}$ because B cannot be observed directly. It was shown in (Blake et al., 1995) that the matrices $\mathbf{A}_0$ and $\mathbf{A}_1$ can be estimated on the basis of the following equations:

$$S_{20} - \hat{A}_0 S_{00} - \hat{A}_1 S_{10} = 0 \tag{10a}$$

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

where $S_{ij} = \sum_{k=1}^{m-2}\left(\mathbf{X}(t_{(k-1)+i})\mathbf{X}^T(t_{(k-1)+j})\right)$, i, j =0,1, 2, and m denotes number of learning frames. Given $\mathbf{A}_0$ and $\mathbf{A}_1$ we can estimate C from the following equation: $\hat{C} = \frac{1}{m-2} Z(\mathbf{A}_0, \mathbf{A}_1)$, where $Z(\mathbf{A}_0, \mathbf{A}_1) = S_{22} + A_1 S_{11} A_1^T + A_0 S_{00} A_0^T - S_{21} A_1 - S_{20} A_0^T + A_1 S_{10} A_0^T - A_1 S_{02} + A_0 S_{01} A_1^T$.

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:

$$\phi_{fg^{(u)}}(t) = (1-\alpha_1)\phi_{fg}(t-1) + \alpha_1 \phi_{fg^{(p)}}(t) \tag{11}$$

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:

$$\phi_{fg}(t) = (1-\alpha_2)\phi_{fg}(t-1) + \alpha_2 \phi_{fg^{(n)}}(t)\,. \tag{12}$$

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

5.1 Experiments in Time-Varying Illumination

To test the elaborated method of skin color segmentation under time-varying illumination we performed various experiments on real images. Some images from one of our test sequences are shown at Fig. 12. Through this sequence we want to highlight the behavior of the tracking algorithm in varying illumination as well as in case of errors in color-based target segmentation. We can notice in frame #56 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 target. In case of such an abrupt change of illumination and without the ASM-based shape refinement the color model would be influenced by the background colors. Thanks to precise delineation of face from the background and the adaptation mechanism the skin model contains only face colors, see frame #60, #70. The accommodation of the skin histograms over time takes place on the basis of 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

illumination and improves tracking performance. In this sequence we can also observe how the size of the shape is scaled in response to varying distance between the moving camera and moving person. The number of learning images has been set to 10.

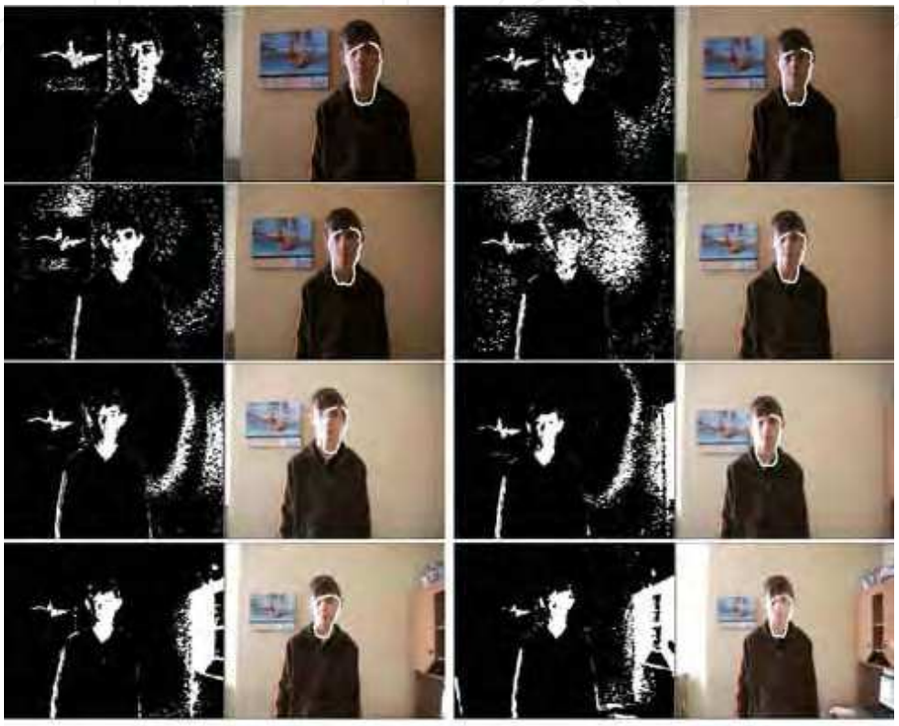


Fig. 12. ASM-based face tracking in varying illumination using intensity gradient and phase of Gabor filter responses. Frames #1, #50, #55, #56, #60, #70, #100 and #200.

To study the adaptation performance in time-varying illumination conditions we conducted experiments with two configurations of the tracking algorithm. In the first configuration only the newly classified pixels were used to accommodate the histogram, whereas in the second one we utilized the predictions of the skin evolution. The predictions lead to better segmenta-tion of the tracked face in varying illumination, see Fig. 13. and Fig. 14. Until significant change of illumination in frame #56, both algorithms produce almost the same results, compare frame #55 at Fig. 13. and Fig. 14. Something better segmentation can be observed as early as in frame #57. Significantly better segmentation can be perceived in frame #70 and all frames behind it. A tracker built on an ellipse can not track the face in frames acquired after the change of the illumination. The presented system runs at 320x240 image resolution at frame rates of 9-11 Hz on a 2.4 GHz PC.

Fig 13. Skin-like regions during adaptation-based on newly classified skin pixels. Frames #55, #56, #57, #58, #59, #60, (top row), #70, #80, #90, #100, #150, #200 (bottom row).

Fig 13. Skin-like regions in learning-based adaptation. Frames #55, #56, #57, #58, #59, #60, (top row), #70, #80, #90, #100, #150, #200 (bottom row).

6. Acknowledgment

This work has been supported by Polish Ministry of Education and Science (MNSzW) within the projects 3 T11C 057 30 and N206 019 31/2664.

7. References

Birchfield, S. (1998). Elliptical Head Tracking Using Intensity Gradients and Color Histograms, In *Proc. IEEE Conf. on Comp. Vision and Patt. Rec.*, Santa Barbara, 232-237

Blake, A., Isard, M., Reynard, D. (1995). Learning to Track the Visual Motion of Contours, *Artificial Intelligence*, Vol. 78, 101-133

Blake, A., Isard, M. (1998). Active Contours, Springer Bradski, G. R. (1998). Computer Vision Face Tracking as a Component of a Perceptual User Interface, In *Proc. IEEE Workshop on Appl. of Comp. Vision*, 214-219

Chen, Y., Rui, Y., Huang, T. (2002). Mode-based Multi-Hypothesis Head Tracking Using Parametric Contours, In *Proc. IEEE Int. Conf. on Aut. Face and Gesture Rec.*, 112-117

Cho, K. M., Jang, J. H., Hong, K. S. (2001). Adaptive Skin Color Filter, *Pattern Recognition*, Vol. 34, No. 5, 1067-1073

Comaniciu, D., Ramesh, V., Meer, P. (2000). Real-Time Tracking of Non-Rigid Objects Using Mean Shift, In *Proc. IEEE Conf. on Comp. Vision and Patt. Rec.*, 142-149

Cootes, T. (2000). 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

Elgammal, A., Duraiswami, R., Davis L. S. (2003). Probabilistic Tracing in Joint Feature-Spatial Spaces, In *Proc. IEEE Conf. on Comp. Vision and Patt. Rec.*, 16-22

Fieguth, P., Terzopoulos, D. (1997). Color-Based Tracking of Heads and Other Mobile Objects at Video Frame Rates, In *Proc. IEEE Conf. on Comp. Vision Patt. Rec.*, 21-27

Han, B., Davis, L. (2005). Robust Observations for Object Tracking, In *Proc. Int. Conf. on Image Processing*, 442-445

Hager, G., Belhumeur, P. (1996). Real-Time Tracking of Image Regions with Changes in Geometry and Illumination, In *Proc. IEEE Conf. on Comp. Vis. and Patt. Rec.*, 403-410

Isard, M., Blake, A. (1996). Contour Tracking by Stochastic Propagation of Conditional Density, *European Conf. on Computer Vision*, Cambridge, 343-356

Koschan, A., Kang, A., Paik, J., Abidi, B., Abidi, M. (2003). Color Active Shape Models for Tracking Non-Rigid Objects, *Pattern Recognition Letters*, Vol. 24, 1751-1765

Kwolek, B. (2006). Active Shape Model-Based Segmentation and Tracking of Facial Regions in Color Images, Int. *Conf. on Image Analysis and Recognition*, Povoa de Varzim, Portugal, Lecture Notes in Computer Science, Vol. 4141, Springer-Verlag, 295-306

La Cascia, M., Sclaroff, S., Athitsos V. (2000). Fast, Reliable Head Tracking under Varying Illumination: An Approach Based on Registration of Texture-Mapped 3D Models, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, 322-336

Nestares, O., Navarro, R., Portilla, J., and Tabernero, A. (1998). Efficient Spatial-Domain Implementation of a Multiscale Image Representation Based on Gabor Functions, *J. of Electronic Imaging*, Vol. 7, 166-173.

Perez, P., Hue, C., Vermaak, J., Gangnet, M. (2002). Color-Based Probabilistic Tracking, *European Conf. on Computer Vision*, 661-675

Raja, Y., McKenna, S. J., Gong, S. (1998). Color Model Selection and Adaptation in Dynamic Scenes, In *Proc. European Conf. on Computer Vision*, 460-474

Soriano, M., Martinkauppi, B., Huovinen, S., Laaksonen, M. (2003). Adaptive Skin Color Modelling Using the Skin Locus for Selecting Training Pixels, *Pattern Recognition*, Vol. 36, 681-690

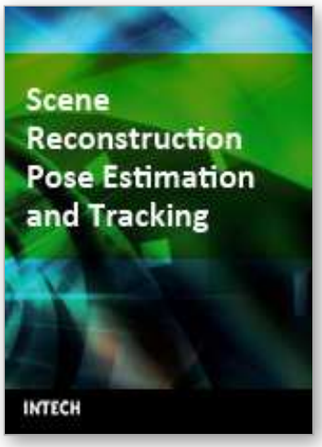
Soriano, M., Martinkauppi, B., Pietikainen M. (2003). Detection of Skin under Changing Illumination: A Comparative Study, *Int. Conf. on Image Analysis and Proc.*, 652-657

Sigal, L., Sclaroff, S., Athitsos, V. (2000). Estimation and Prediction of Evolving Color Distributions for Skin Segmentation under Varying Illumination, In *Proc. IEEE Conf. on Comp. Vision and Patt. Rec.*, 2152-2159

Sobottka, K., Pitas, I. (1996). Segmentation and Tracking of Faces in Color Images, In *Proc. 2-nd Int. Conf. on Automatic Face and Gesture Rec.*, 236-241

Srisuk, S., Kurutach, W., Limpitikeat, K. (2001). A Novel Approach for Robust Fast and Accurate Face Detection, *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 9, No. 6, 769-779

Swain, M. J., Ballard, D. H. (1991). Color Indexing, Int. J. of Comp. Vision, Vol. 7, No. 1, 11-32 Yang, J., Weier, L., Waibel, A. (1998). Skin-Color Modelling in Color Images, In Proc. *Asian Conf. on Computer Vision*, II:687-694



Scene Reconstruction Pose Estimation and Tracking

Edited by Rustam Stolkin

ISBN 978-3-902613-06-6

Hard cover, 530 pages

Publisher I-Tech Education and Publishing

Published online 01, June, 2007

Published in print edition June, 2007

This book reports recent advances in the use of pattern recognition techniques for computer and robot vision. The sciences of pattern recognition and computational vision have been inextricably intertwined since their early days, some four decades ago with the emergence of fast digital computing. All computer vision techniques could be regarded as a form of pattern recognition, in the broadest sense of the term. Conversely, if one looks through the contents of a typical international pattern recognition conference proceedings, it appears that the large majority (perhaps 70-80%) of all pattern recognition papers are concerned with the analysis of images. In particular, these sciences overlap in areas of low level vision such as segmentation, edge detection and other kinds of feature extraction and region identification, which are the focus of this book.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen