# Real-Time Model-Based Hand Localization for Unsupervised Palmar Image Acquisition

Ivan Fratric[1], Slobodan Ribaric[1]

[1] University of Zagreb, Faculty of Electrical Engineering and Computing,
Unska 3, 10000 Zagreb, Croatia
{Ivan.Fratric, Slobodan.Ribaric}@fer.hr

**Abstract.** Unsupervised and touchless image acquisition are two problems that have recently emerged in biometric systems based on hand features. We have developed a real-time model-based hand localization system for palmar image acquisition and ROI extraction. The system operates on video sequences and produces a set of palmprint regions of interest (ROIs) for each sequence. Hand candidates are first located using Viola-Jones approach and then the best candidate is selected using model-fitting approach. Experimental results demonstrate the feasibility of the system for unsupervised palmar image acquisition in terms of speed and localization accuracy.

**Keywords:** Hand localization, real-time, model-based, unsupervised biometrics.

## 1 Introduction

In biometric systems based on hand features, two problems have recently emerged: unsupervised and touchless image acquisition. In previously developed systems [1 - 8], the image is taken under strictly controlled (light and position) conditions. In most of the systems the user is required to place a hand on the sensor. However, this is a problem for large-scale scenarios because many people refuse to touch the biometric sensor for sanitary and other reasons. Also, touchless hand-based biometric systems in combination with unsupervised image acquisition are more convenient to use.

The motivation of our work is the development of a touchless palmprint recognition system. The system should aslo be unsupervised in the sense that there is no need to guide a user during the acquisition procedure; it is sufficient for the user to wave the hand in front of a sensor or a group of space-distributed sensors.

The system should be robust enough to be able to work in uncontrolled environment conditions with varying lighting and a cluttered background, as required by most real-world applications. The robust real-time detection and localization of the hand are crucial in order to be able to develop such a system.

In this paper we describe the development of a real-time model-based hand localization system for palmar image acquisition. The system operates on video sequences taken in a real environment and produces a sequence of images containing the palmprint regions of interest (ROIs). This sequence of ROIs is intended to be the input for a biometric verification system, which would select the best ROIs in terms of feature extraction and use them for transparent user verification.

Scanner-based hand biometric systems have been proposed in [1, 2, 3]. However, scanners are slow and require the hand to be placed on the scanning surface. Zhang et al. [4] developed an online palmprint identification system in which a hand image is captured using a CCD camera, but the user is still required to place a hand on the device and pegs are used to constrain the position of the hand. Papers [5, 6] describe biometric systems based on the hand's geometry. Both systems use cameras as the input devices and require the user to place the hand on the surface, where it is constrained by pegs. The multimodal biometric systems described in [7, 8] use a camera-based input device and do not restrain the position of the hand, but the hand is still required to be placed on a flat, uniformly colored surface.

In a recently published paper, Ong et al. [9] describe a touchless palmprint verification system. Skin-color-based segmentation is used in order to segment the hand in the image, which may not work in environments with skin-colored backgrounds or poor lighting. Their system is set to extract a single ROI every 2 seconds, which is then used for the verification.

Most of the work related to hand detection in real environments is in the area of natural human-computer interaction. A review of techniques used in this area is given in [10]. Many of the techniques described use motion information and/or heuristical approaches for hand detection. However, in our scenario there could be multiple moving objects on the image, so this information alone would not be sufficient for robust hand detection.

Little work has been done on finding hands in images based on their appearance and shape. Kölsch and Turk [11] used an object detector proposed by Viola and Jones to locate hands. Their approach is fast, but finds only hands in a pre-defined pose and is unable to determine the contour on the hand in the image. Stenger et al. [12] use a hierarchical model-based approach for finding and tracking hands. Their approach finds the hand shape and orientation, but takes approximately 2 seconds per frame to execute on a 1-GHz Pentium IV.

## 2 System Organization

Fig. 1 shows the overview of the developed hand localization system for palmar image acquisition and ROI extraction.

A camera is used to capture a video sequence of the hand at a resolution of 640x480 pixels, which is a trade-off between image quality and processing time. Each of the video frames is then processed as follows.

Firstly, the frame is scaled to a resolution of 320x240 pixels in order to decrease the time required for the image processing. This reduced image (converted into grayscale with 256 levels) is used for the detection of the hand candidates. The Viola-Jones approach is used for hand-candidate detection because of its speed and good detection rate [13].

Once the positions of the hand candidates have been located on the image, each of the candidates is verified by fitting it to the large, predefined set of hand models. Two measures are used for the correspondence of the hand candidate and the model [12]: the edge distance and the skin-color likelihood. These two measures are combined in a unique measure that determines which of the candidates, if any, should be used for the feature extraction.

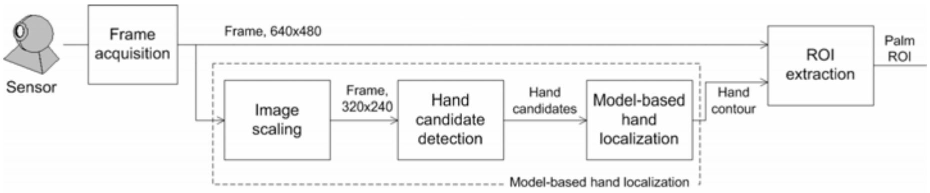

**Fig. 1.** Overview of the hand localization system for palmar image acquisition and ROI extraction

The output of the model-based hand localization stage is a hand contour with marked stable points, which are used in the ROI extraction stage to determine the position of the palm ROI on the original (non-scaled) image.

## 3 Hand-Candidate Detection

The goal of the hand-candidate detection process is to obtain a relatively small set of image locations where the presence of the hand is possible. This process should be as fast as possible so as to be able to process the entire frame in real time.

The Viola-Jones approach was selected for the candidate detection [13]. This approach has become very popular in computer vision because of its high speed and good accuracy. It has been primarily used for face detection [13]; however, it has also been used for hand detection [11].

The Viola-Jones object detector operates on grayscale images. A sliding-window approach is used to examine each image location for the presence of a hand. The classifier itself is built as a cascade of weak classifiers. These weak classifiers are constructed using AdaBoost on a large set of 2D Haar-like features, which can be computed quickly by using integral images (each image element on the integral image contains the sum of all the pixels to its upper left on the original image). Weak classifiers are trained to have a very high detection rate (99.5% in our case) and a high false positive rate (50% in our case). This means that at each stage of the classification most of the real hands will be passed to the subsequent classifiers in the cascade, while a part of the non-hands will be rejected. Thus, most of the non-hands will be rejected in the first several stages of the cascade, which, together with fast feature computation, makes real-time object detection possible.

Usually, after the detection, the results are filtered by grouping overlapping detection windows and eliminating groups with a small number of members. However, we skipped this step because we wanted to pass all of the detection to the model-fitting stage in order to achieve the best possible fitting.

We used the OpenCV [14] implementation of the Viola-Jones object detector.

## 4 Model-Based Hand Localization

In the model-based hand localization phase, each of the candidates for the hand position, found in the previous stage, were examined by fitting a large number of hand models over that position on the image. The goal of this process is to further verify the hand candidates, remove the non-hands, and only select the candidate with the highest matching score. The output of this stage is an approximate hand contour – that of the best-fitting model.

The model-matching process is taken from [12]. Each model is represented as (a) a contour, (b) a binary mask, where white points indicate the inside of the hand and black indicate the outside of the hand.

Unlike [12], where multiple hand models were obtained by rotating a small set of models covering basic hand gestures, in our system models are generated from the database of 1872 hand images of 243 people, obtained using a table scanner (180 dpi). The hands in the database images are cropped and rescaled to the resolution of 64x64 pixels to form the models. This set of models contains significant variations in hand positions and rotations. Fig. 2 shows some of the models used in our system.

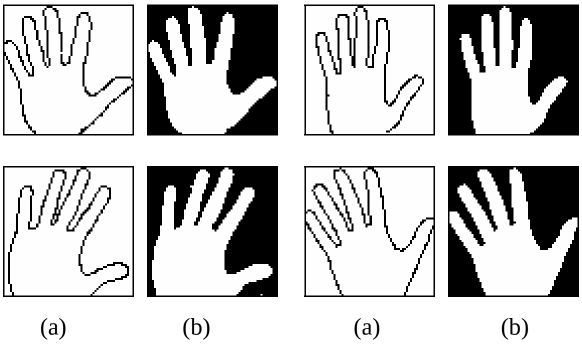

**Fig. 2.** Some of the hand models used in the model-fitting stage: (a) the hand contours, (b) the corresponding binary masks

Two measures are used for the correspondence between the model and the hand candidate: the edge distance and the skin-color likelihood. The edge distance is computed as the quadratic chamfer distance between the model contour and the edges of the image. For two sets of points, *A* and *B*, it is given as

$$d(A,B) = \frac{1}{N_a} \sum_{a \in A} \min_{b \in B} \|a - b\|^2 \qquad (1)$$

where $N_a$ denotes the number of points in the set *A*. In our system *A* is the set of points on the model contour, while *B* is the set of the images' edge elements. The edges on an image are obtained using the Canny edge detector [15]. The computation speed of this distance can be increased if a distance transform of the edge image is computed prior to the model-fitting stage.

The hand's skin-color likelihood, given the model x, is computed as

$$p(hand \mid Model\ x) = \prod_{k \in S(Model\ x)} p^{s}(I(k)) \prod_{k \in \bar{S}(Model\ x)} p^{bg}(I(k)) \qquad (2)$$

where $S(Model\ x)$ denotes the set of points inside the hand in the model *x*, where $\bar{S}(Model\,x)$ denotes the set of points outside the hand in the model *x*, *I(k)* is the color of the image at location *k* and $p^{s}$ and $p^{bg}$ are the skin and the background color distributions.

The skin and the background color distributions are taken from [16]. A non-parametric color model is used, where the skin and background color distributions are computed based on RGB histograms with R, G and B resolutions of 32x32x32 obtained from a large database of images with manually marked skin regions.

This likelihood in (2) is easier to compute as a log likelihood

$$\log(p(hand \mid Model\,x)) = \sum_{k \in S(Model\,x)} \log(p^{s}(I(k))) + \sum_{k \in \bar{S}(Model\,x)} (\log p^{bg}(I(k))) \qquad (3)$$

The hand matching score between the model and a hand candidate is obtained by first normalizing the edge distance and skin-color likelihood using the min-max normalization [17] and then summing the two normalized scores.

If speed was not one of the requirements of the system, the process of model fitting for a hand candidate would involve iteratively matching all the models to the hand candidate and selecting the best-fitting model based on the matching score. However, this process would be very slow, so instead, in our system, each hand candidate is matched only to the subset of models arranged as a 3-level tree. This 3-level tree contains all 1872 models, in which each node contains a group of models (and its prototype). Each child-node in the tree is obtained by grouping models contained in its parent node based on a modified simple heuristical clustering algorithm. The nodes at the third level represent individual models.

A modified simple heuristical clustering algorithm is given below:

```
Set the first model as the prototype for the first cluster
For each model m
 |    If the distance between model m and any prototype
 |    is more than characteristic distance
 |     |    Create a new cluster and set m as its
 |     |    prototype;
For each model m
 |    For each cluster C
 |      |    If the distance between the model m and the
 |      |    prototype of C is less than characteristic
 |      |    distance
 |      |      |    Put m in C;
```

Note that in the above algorithm one model can end up in different clusters. The edge distance between the model contours computed according to (1) is used as the measure of the distance between models. The characteristic distance is selected experimentally as 8 at the level 1 of the tree and 4 at the level 2 of the tree.
Once the tree is constructed, the fitting process for each candidate can be performed as follows

```
1. Set current node to tree root
2. Compare hand candidate to the prototype of every
   child of the current node
3. Set current node to the child with the highest hand
   matching score
4. If current node has children go to step 2.
5. Best matching model is the one contained in the
   current node
```

This process is repeated for all the hand candidates and the best candidate in terms of the hand matching score is selected. This hand matching score is then compared to the threshold T, set to 1 (50% of the maximum expected hand matching-score value), which was selected based on preliminary experiments on the training set. The purpose of this thresholding is to eliminate the non-hands and also hands with strong shape deformations (e.g. closed or half-closed hand). If the score is lower than T, it is assumed that the matching is not good enough and there is no hand in the current frame. Otherwise, the contour of the best-matching model is assumed to be the contour of the hand on the image.

We also experimented with using snakes [18], initially positioned as the contour of the best-matching model, to further adapt the contour to the hand in the image. However, it was observed that this step was not robust enough to be used in environments with a cluttered background, and that better fitting is achieved if it is skipped.

## 5 ROI Set Extraction

The inputs to the ROI extraction module are: (a) the hand contour of the best model with marked stable points, and (b) the original frame (640x480 pixels).

A simplified procedure described in [3] was used for determining stable points on the contour of each model (Fig. 3). Based on these stable points, the location of the palm ROI in the current frame is determined.

In our system the palm ROI is defined as a square region with two of its corners placed on the middle-points of the line segments P1-B2 and B4-P2 (see Fig. 3).

This region is extracted from the current frame and geometry normalization is applied in order to obtain a ROI image of fixed size and orientation.

Any additional normalization, such as lighting normalization, can be applied at this stage as well.

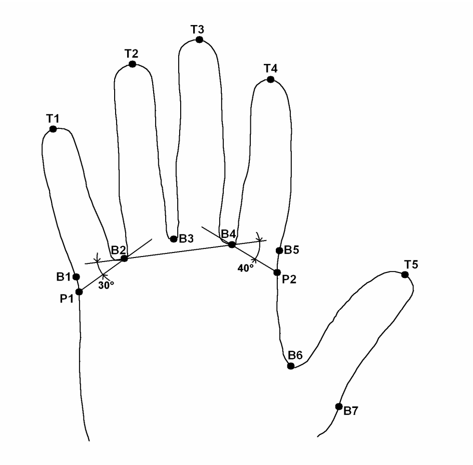

**Fig. 3.** Hand contour with the stable points marked

## 6 Experimental Results

The experiments were conducted on a database of video sequences. The database was divided into a training set and a test set. The training set consists of 32 video sequences of 10 people. A total of 2138 frames containing hands were extracted and used for the training. The test set consists of 10 video sequences of 5 people. The people in the test set are not the same as the people in the training set. A total of 362 frames were randomly selected for testing from the test sequences.

The database was collected using two cameras: a high-quality CCD camera and a low-quality CMOS web camera. The sequences were captured under varying lighting conditions and all of them feature a cluttered background.

We first tested the detection using the Viola-Jones hand detector only, and then using the full model-based detector. The results of the detection using only the Viola-Jones object detector can be seen in the ROC curve (detail) in Fig. 4.

The results are shown for two sizes of detection window. As can be seen from the ROC curve, high detection rates can be achieved using the Viola-Jones detector. Because the results for different sizes of detection windows were similar in performance, a smaller window of 20x20 pixels was selected for further experiments. A detection rate of 98.3% can be achieved with 19 false positives, and a detection rate of 99.7% can be achieved with 713 false positives. This point in the ROC curve was selected as the working point of our system. This maximizes the detection rate, and the false positives can be eliminated in the model-fitting stage.

In the second experiment the model-based localization and ROI extraction procedure, as given in the previous sections, were applied to all the frames in the test set. Fig. 5 shows some of the successful and non-successful ROI localization results.

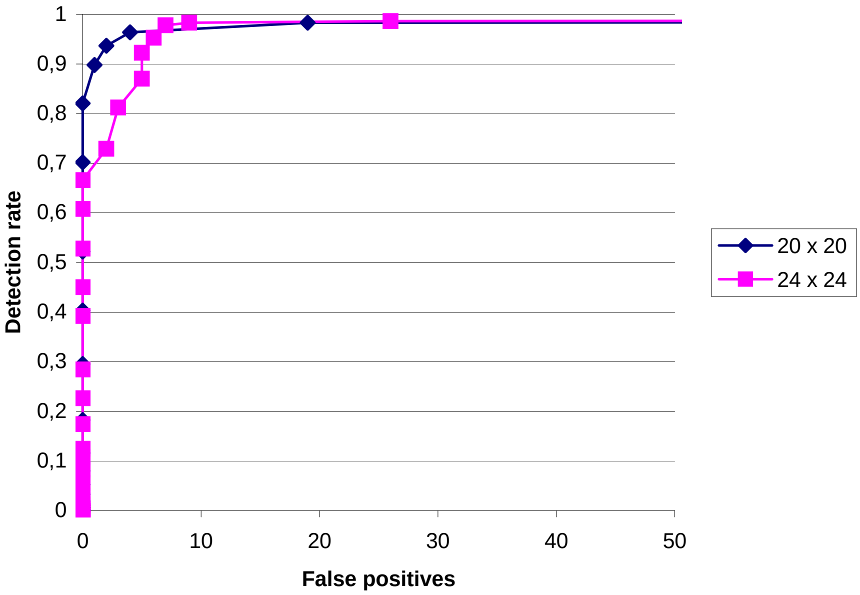

**Fig. 4.** ROC curve for hand detection using the Viola-Jones object detector

As can be seen from Fig. 5, the system can operate successfully in scenes with a cluttered background, even if the hand is placed over other skin-colored regions, such as the face. The errors are mostly due to skin-colored regions with edge elements resembling, in terms of shape, edges typically found on a hand, or a hand being incorrectly positioned on the image.

To measure the accuracy of the system, we defined the successful ROI localization as localization where the ROI falls inside the palm on the actual frame. On our test set, with the accuracy measure as described above, the system accurately locates the palm ROI in 96.9% of cases.

The average processing time for a single frame on a single processor with 1177 SPECint_base2000 was 0.27 seconds. However, hand detection in the verification system is intended to be run on a newer, quad-core processor, where each core would be assigned one frame. In this setup, taking into account the ability of the detection system to independently process frames, the average processing time could be reduced by up to four times.

## 7 Conclusion

We have developed a hand-localization system for palmar image acquisition and ROI extraction. The system first locates the hand candidates using the Viola-Jones approach and then selects the best candidate using a model-fitting approach.

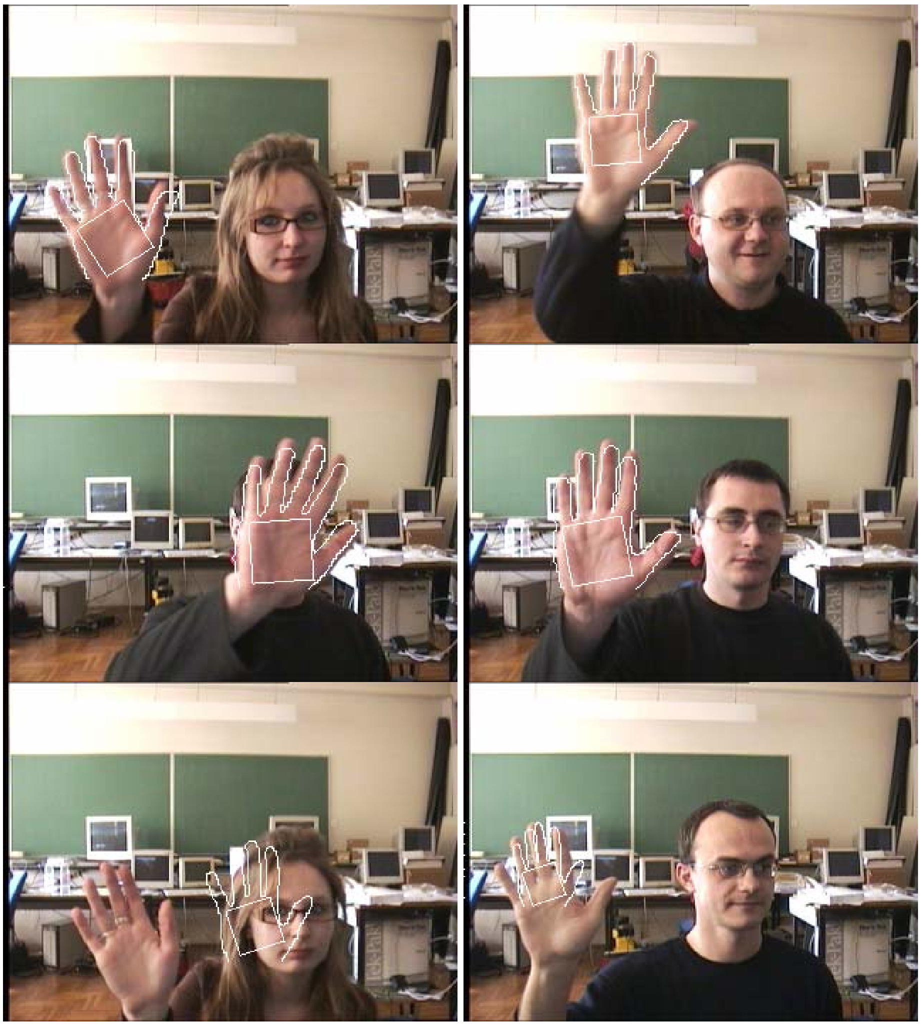

**Fig. 5.** Some of the results of the hand and ROI localization (successful and unsuccessful)

The Viola-Jones approach is very fast and highly accurate for hand detection (a 98.3% detection rate with 19 false positives). The model-fitting process selects the best candidate and finds a palm ROI that falls inside the palm on the image in almost all frames (96.9% in our test set). This demonstrates the feasibility of our system, even in environments with cluttered or skin-colored backgrounds. The system is designed to operate in real-time on a multi-core processor.

In the future we plan to use this system as part of a touchless palmprint verification system. To avoid any contact, the identity could be presented to the system by RFID or a similar unsupervised technology. The unsupervised characteristic of the system could be achieved by using a group of space-distributed (intelligent) sensors.

The system would be operating on better equipment than that used in the experiments described in this paper, such as a high-speed multi-core processor together with a high-quality camera, which would enable us to obtain high-quality features usable for the identity verification. Techniques for selecting the best ROIs for verification will be developed and tested in terms of the FAR and the FRR.

## References

1. Han, C., Cheng, H.L., Fan, K.C., Lin, C.L.: Personal Authentication Using Palm-print Features. Pattern Recognition, Vol. 36, pp. 371-381 (2003)
2. Lin, L., Chuang, T.C., Fan, K.C.: Palmprint verification using hierarchical decomposition. Pattern Recognition Vol. 38, No.12, pp. 2639–2652 (2005)
3. Ribaric, S., Fratric, I.: A Biometric Identification System Based on Eigenpalm and. Eigenfinger Features. IEEE Trans. PAMI, Vol. 27, No. 11, pp.1698-1709 (2005)
4. Zhang, D., Kong, W.K., You, J., Wong, M.: Online Palm Print Identification. IEEE Trans. PAMI, Vol. 25, No. 2, pp. 1041-1050 (2003)
5. Jain, A.K., Ross, A., Pankanti, S.: A prototype hand geometry-based verification system. Proc. 2nd Intl. Conf. on Audio- and Video-Based Biometric Person Authentication, pp. 166–171, Washington DC, USA (1999)
6. Sanchez-Reillo, R., Sanchez-Avila, Gonzalez-Marcos, A.: Biometric identification through hand geometry measurements. IEEE Trans. PAMI, Vol. 22, No. 10, pp. 1168–1171 (2000)
7. Kumar, A., Wong, D.C.M., Shen, H., Jain, A.K.: Personal verification using palmprint and hand geometry biometric. Proc. Intl. Conf. Audio- and Video-based Person Authentication, pp. 668-675 (2003)
8. Kumar, A., Zhang, D.: Personal Authentication using Multiple Palmprint Representation. Pattern Recognition, Vol. 38, No. 10, pp 1125-1129 (2005)
9. Ong, M.K.G., Connie, T., Teoh, A.B.J., Touch-less palm print biometrics: Novel design and implementation. Image and Vision Computing, Vol. 26, No. 12, pp.1551-1560 (2008)
10. Erol, A., Bebis, G., Nicolescu, M., Boyle, R., Twombly, X.: A Review on Vision-Based Full DOF Hand Motion Estimation. Proc. IEEE Workshop on Vision for Human-Computer Interaction, pp 75-82, USA (2005)
11. Kölsch, M., Turk, M.: Robust Hand Detection. Proc. IEEE Intl. Conference on Automatic Face and Gesture Recognition, pp. 614-619 (2004)
12. Stenger, B., Thayananthan, A., Torr, P.H.S., Cipolla, R.: Model-based hand tracking using a hierarchical Bayesian filter. IEEE Trans. PAMI, Vol. 28, No. 9, pp. 1372–1384 (2006)
13. Viola, P., Jones, M.: Robust Real-Time Object Detection. Intl. Journal of Computer Vision, Vol. 57, No. 2 (2004)
14. Open Computer Vision Library, http://sourceforge.net/projects/opencvlibrary/
15. Canny, J.: A Computational Approach To Edge Detection. IEEE Trans. PAMI, Vol. 8, pp. 679-714 (1986)
16. Jones, M.J., Rehg, J.M.: Statistical color models with application to skin detection. Intl. Journal of Computer Vision, Vol. 46, No. 1, pp. 81-96 (2002)
17. Jain, A.K., Nandakumar, K., Ross, A.: Score Normalization in Multimodal Biometric Systems. Pattern Recognition, Vol. 38, No. 12, pp. 2270-2285 (2005)
18. Kass, M., Witkin, A., Terzopoulos, D., Snakes: active contour models. Intl Journal of Computer Vision, Vol. 1, No. 4, pp. 259–268 (1988)