Hand detection for contactless biometrics identification

Julien Doublet, Olivier Lepetit, Marinette Revenu

Hand Detection for Contact less Biometrics Identification

Julien Doublet[1], Olivier Lepetit[1], Marinette Revenu[2]

1: France Telecom, 42 rue des coutures, B.P. 6243, 14066 CAEN Cedex 4, France
2: GREYC – ENSICAEN, 6 Bd Maréchal Juin, 14050 CAEN Cedex, France

Abstract: In this paper, we present a new hand detection system based on a combination of skin color modelling and active shape model. Skin color distribution specified by a multiresolution neural network constrains active shape model evolutions on skin objects. We present the skin color detection performances of this near real time system on many color spaces and show that a combination of skin color and active shape model increase robustness and accuracy of hand detection in complex images.

Keywords: Image Processing, Skin Detection, Active Shape Model, Machine Learning, Biometrics.

1. Introduction

Face and hand detection has been largely studied these last years. It constitutes, indeed, the first stage of gravitational applications energy of biometrics recognition to man-machine interaction.

Detection processes are multiples. For face detection [1], methods as neural networks (NN) [2], Markov models [3], generic models [4] or more recently convolutional neural networks [5] are used to extract information contained in face. For hand segmentation, detection is articulated on contour methods using active contours [6] or, for a more robust approach, active appearance models (AAM) or active shape models (ASM) [7].

However, skin color segmentation remains one of the methods the most employed for face and hand detection. Although not very robust to important luminosity changes, it can according to application nature be used alone, or thanks to its weak computational cost, like filtering in order to improve segmentation speed of more expensive methods. To limit problems of classical processes, a combination of accurate skin color segmentation and robust active shape model is presented for hand detection.

The paper is organized as follows: section 2 presents our method on pixel-based skin detection technique. In Section 3, a modified active shape model is described. The final algorithm is presented in section 4. Experimental results and conclusion are given in section 5 and 6, respectively.

2. Skin Color Segmentation

The aim of skin color modelling is to classify color pixels like skin or not, without using neighbourhood information.

2.1 Skin color detection

Classical skin distribution models are defined assuming some restrictions. In one hand, models representing skin color without generalisation, like Lookup table methods [8] or Bayes classifiers [9], depend on database quality. It must completely describe all skin and background tones, that it is non-realistic in practice. In the other hand, parametric models like Gaussian Mixture Model [10] suppose skin color distribution can be mathematically modelling.

Without assuming these constraints, machine learning can model skin tone. Thanks to its good estimation process and quick execution, neural networks [11] introduced by McCulloch and Pitts are one of the most used machine learning. Three neural network configurations are used to model skin tone. The two first contain one hidden layer with $N \in [1,10]$ neurons. The first one has a sigmoid transfer function while the second one has a hyperbolic tangent transfer function. The third configuration is a network with two hidden layers. The first layer contains $N1 \in [1,10]$ neurons with a sigmoid transfer function; the second one contains $N2 \in [1,10]$ neurons with a hyperbolic tangent transfer function. Some another configurations with more neurons or more hidden layers aren't used to limit computations (cf. 5.2). The inputs of each network are the pixel's color components in RGB, YUV, HSL and/or TSL color domain. The output is composed by one neuron whose the value $s \in [0,1]$ corresponds to the probability that a color is a skin color.

In learning phase, an image database is created and color balanced to resist on luminosity change. Next, all pixels are marked skin or no skin. The neural network is trained with the output value equals to one when a color is a skin color and zero otherwise. In evaluation phase, the image is color balanced and the probability that each pixel in image is a skin pixel is computed by the neural network (fig. 1). This complete process builds the probability map of the input image.

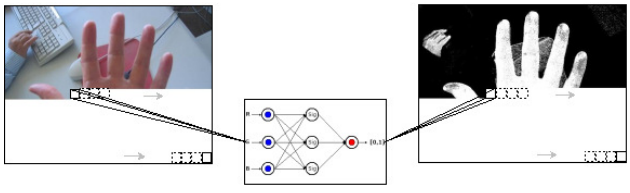

Figure 1 : Probability map computation

Although the neural network is prompt to compute a probability map (cf. 5.2), for a near real time skin detection, a pyramidal algorithm is used. This method is based on iterative treatments of the input image according to his resolution rate (fig 2) and it can be explained by:

- Compression of the original image to resolution K_i
- Computing of the probability map PM_i
- Extrapolation of PM_{i+1} to resolution $K_{i+1} > K_i$
- For each pixel to a distance $d<D$ from skin/no skin variation, re-computing of probability that this pixel is a skin pixel at this resolution. D denotes the exactness of the computation
- Back to extrapolation until $K_i=C$ where C denotes the final resolution

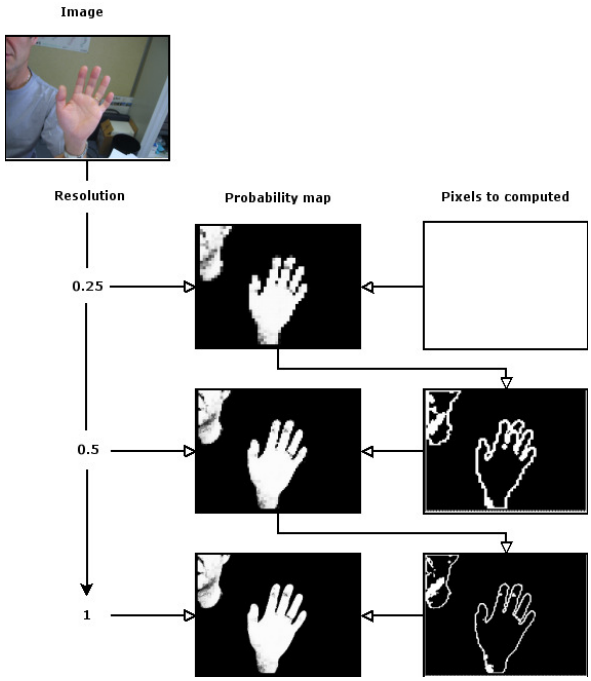

Figure 2 : Pyramidal probability map computed with K_0=0.25, K_1=0.5, K_2=C=1,D=1

2.2 Post-processing

From the probability map, to extract hand is a complex process. Many different methods are used to identify the detected skin object (hand or face). The principal method consists on segmentation by histogram thresholding on probability map [12] [13]. Other methods are defined: Brèthes and al. [14] use watershed algorithm, Lui and Izquierdo [15] use median filtering while Tomaz and al. apply a skin pixels grouping [16].

Our approach is made up a morphological dilatation with a circular element of radius one pixel. Next, the probability map is segmented by histogram thresholding. The threshold is fixed to 0.5; each pixel is considered like skin color if the neural network output value *s* is more that 0.5 else it is classified like background tone.

The skin detection can't separate two skin tone objects. In order to distinguish them, the image contours calculated by Di Zenzo algorithm (DZA) [17] in skin color space is used. This skin color space is computed using Principal Components Analysis (PCA) [18] on the skin pixels database in RGB space. From contours, in assuming that gradient between two objects is highter than a fixed threshold δ, all skin pixels whose gradient is highter than δ are regarded as background pixels. Finally, as our hand detection process will be used for biometrics recognition, we suppose that only one hand is presented to the system, so only the region with the largest surface is preserved. This complete process is illustrated in figure3.

Figure 3 : Post-processing for hand segmentation (a) probability map (b) probability map thresholded (c) gradient difference and region selection

3. Color active shape model

The active shape model, introduced by Tim Cootes [7] to segment medical images, uses a shape model, the Point Distribution Model, to detect objects of a known shape in complex images.

3.1 Point Distribution Model (PDM) computation

Skin color segmentation is not sufficient to robustly and precisely detect a hand in complex images. Two major difficulties cannot be completely solved: luminosity change influences skin color detection quality and it's difficult to directly differentiate a hand from a face from segmented image. Active shape model can solve these two unsatisfactory problems. It resists on luminosity change and the hand shape learning makes it possible to differentiate a hand from a face.

In our active shape model, the landmark points representing hand shape and allowing to compute the PDM are automatically determined from segmented hand. Each hand in a learning database is represented by a vector X with n landmarks points Xi:

$$\mathrm{X} = (X_1, \ldots, X_n)^T \qquad [1]$$

These landmarks are defined using hand shape characteristics: hand contour contains nine geometrical landmarks – the fingertip points and the valley points between adjacent fingers (fig. 4). As the landmark points determination will be used after a no controlled skin color segmentation to init active shape model too, the contour could be not regular. So,the landmarks points aren't extracted by travelling along hand boundary by contour curvature analysis as in [19]. They are computed by detecting the five maxima and four minima in a distance diagram corresponding to distance between each contour points and a fixed point (fig 4). The other points describing hand shape are fixed at an equidistant distance between its landmark points (fig 4).

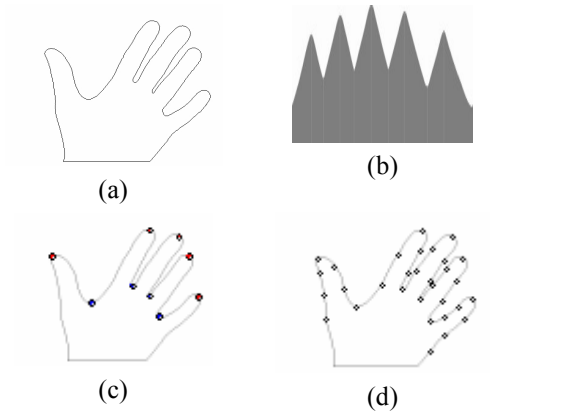

Figure 4: Automatic landmarks extraction (a) hand shape (b) distance diagram (c) characteristic points (d) anoted hand

3.2 Search method

The research phase in ASM supposes a good initialisation. The initial hand shape must be close to the real hand shape in the image, that's a difficult problem. Some general global optimisation techniques, such as Genetic Algorithms [20] can be used to this initialisation. But, for a faster algorithm, the model is initialized from skin color segmentation. The landmarks are computed by the same process as the learning phase.

In searching phase, a default of active shape model is the proportion the model shape increases without followed real hand contour. Its mistakes happened when intialisation isn't enough accurate or when some strong gradients are around the object to detect. Two combined methods are used to remove this phenomena. In the first time, the image gradients *G* are determined using DZA in skin color space that limit intensity gradients in background. In second time, the probability *s* that a pixel is skin colored is used to adjust the gradients *Ga* (fig. 5):

$$G_a = G \times s \qquad [2]$$

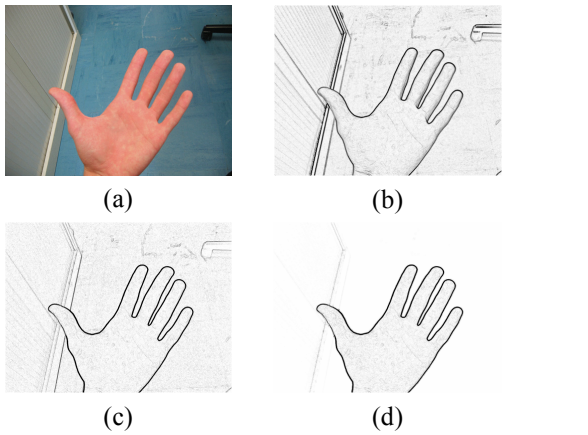

Figure 5: Gradient determination (a) original image (b) DZA in RGB space (c) DZA in skin space (d) ajusted DZA in skin space

The model's points are deformed in this gradient image that limits contour divergences.

For more expansion limitation and for final shape correponds to real hand contour, a weight $W \in [0,1]$ is applied to deformations. When the weight is equal to one, only ASM constraints are applied, when W is equal to zero, the points are deplaced without constraints following the normal contour (like active contour models).

4. Final Algorithm

The final algorithm can be described into the following steps:

- Color balancing of the image.
- Computation of the probability map by pyramidal algorithm.
- Post-processing of the probability map.
- Initialisation of the active shape model.
- Computation of the adjusted gradients.
- Deformation of the model following the PDM constraints and the adjusted gradients.

5. Experimental testing

In this section, experimental results are indicated for the parameters determination and to validate our approach. First, the experimental databases are described in section 5.1. Next, the skin color detection results and the quality evaluation of the complete process are demonstrated in section 5.2 and 5.3, respectively.

5.1 The database

All images used to validate the methods come from internet or our acquisitions. The images are split into two sets: the skin detection set and the ASM set.

The skin detection set is divided in two sets. The first one, corresponding to the learning set, contains 5789766 pixels of which 2189796 skin color pixels. The second one for tested skin detection contains 607800 pixels equally distributed among skin and no skin pixels.

The ASM set is composed of 260 images of size 454*341. It is split into three sets. The first set contains 86 segmented hand images whose the landmarks points are automatically extracted. The second set, which permits to adjust constants, contains 87 marked images. The last set contains 87 marked images too in order to validate the complete process.

5.2 Color model Learning

Bayes' estimation, Gaussian Mixture Model and Neural Network approach are compared on the skin detection set. The table 1 and the table 2 show skin segmentation rate without post-processing when the false positive rate and false negative rate are equals. The false negatives correspond to background pixels classify like skin pixel, the false positives to skin pixels classify like background.

Domain	Bayes
RGB	13.6%
YUV	**9.4%**
UV	12.1%
HSL	12.9%
HS	11.3%
TSL	11.6%
TS	12.6%
RGBn	11.9%

(a)

Gaussians	UV	HS
1	19.0%	33.4%
2	19.2%	18.6%
3	21.1%	15.6%
4	16.9%	15.7%
5	17.5%	15.4%
6	17.0%	15.3%
7	**16.7%**	**15.1%**
8	17.3%	15.6%

(b)

Table 1: Error rate with Bayes' estimation (a) and Gaussian Mixture Model (b) for skin detection

It is interesting to remark that majority of articles don't use the luminosity component in YUV, HSL and TSL domain which has, in fact, an important impact in error rate. Compare to Bayes' method, Gaussian (Mixture) Model presents an error rate most important which can be explain by an excessive generalisation.

The table 2 shows segmentation rate with neural network according to different color domains and network architectures. With this method, the error rate is smaller than for Bayes' estimation and Gaussian Mixture Model. The skin representation complexity can explained this result. Nevertheless the hidden layers additions or many color representations in neural network input don't fundamentally influence the error rate.

Domain	Sigmoid	Tanh	Sigmoid+Tanh
RGB	**6.2%**	7.8%	6.8%
YUV	6.7%	7.2%	6.4%
HSL	7.9%	8.1%	7.0%
TSL	6.8%	7.0%	**6.2%**
RGB+YUV	7.3%	7.3%	6.8%
RGB+TSL	**6.2%**	**6.2%**	6.3%
YUV+TSL	6.7%	**6.2%**	6.4%

Table 2: Error rate with neural networks for skin detection

To complete this evaluation, the neural network approach has been tested in skin color space and compared to Support Vector Machine introduced by Vapnik [21]. The skin color space has been tested with its two first components (PCA_1) and the two last (PCA_2). The error rate is described in table 3.

Domain	Sigmoid	SVM
RGB	6.2%	**4.9%**
PCA_1	12.9%	**12.0%**
PCA_2	11.8%	**11.0%**

Table 3: Error rate in skin color and RGB space with neural network and SVM

The error rate in RGB space increases of more 20% between neural network and SVM. Otherwise, skin color spaces give a similar rate than others domains without luminance information.

The execution time is a dominating factor in segmentation process. The average computing time of the major approaches is indicated in the table 4 where P(x) is a pyramidal approach applied to detection method X. For this experiment, the SVM, being more expensive than NN, can't be used despite its better segmentation rate for a near real time approach.

Domain	NN	P(NN)	SVM	P(SVM)
RGB	2624ms	155ms	69097ms	4237ms

Table 4: average computing time for images of size 454*341

5.3 Point Distribution Model and Detection evaluation

The shape model is defined by the Point Distribution Model computed with the first ASM set. In saving 98% of informations, only the first 13 axis (eigenvectors) of 222 are kept to model hand shape. The figure 9 illustrates hand's variations in the first three axis with a limited deformation. The landmarks points being automaticly located, the wrist's points are unaccuratly detected. Netherless, our goal is to detect the fingers' shapes for biometrics recognition, so the wrist's points location are not really important here. For accurate definition of these points, a hand landmarking of the learning set can be defined.

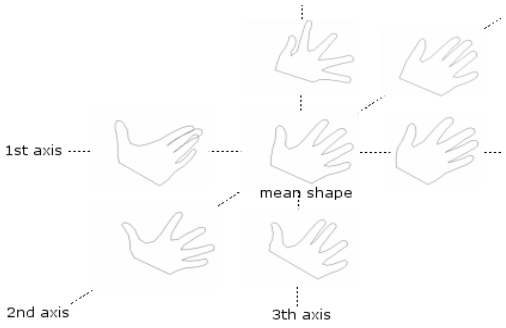

Figure 6: Hand shape variations around mean shape according to the third principal axis

All constants must be fixed to evaluate detection quality of the complete process. First, the plans of skin color space to use and the threshold δ are determined by experimentations.

The effectiveness of gradient difference approach is calculated by the ratio between the average of intensity gradients in a zone close to the real hand's contour and in the image remainder. The results of its study, described in table 5, show that the second plan is the most discriminant plan between all the combinations.

plan	1st	2nd	3rd	1&2	1&3	2&3	All
Ratio	3.46	2.22	2.41	3.59	3.33	3.55	3.49

Table 5: Ratio between the average of intensity gradients of the hand's contour and the background

We defined an error distance to compare two shapes, and so to evaluate quality detection. This error distance is the average of minimum euclidian distance between each landmarks points X_i and the ideal contour C:

$$e = \frac{1}{N} \sum_{i=1}^{N} \min(\ d(X_i, C)) \qquad [3]$$

Where d(a,b) is the Euclidian distance between the points a and b. This error distance e is computed with different δ values to find the optimal parameter. After this experimentation, the lower error distance is determined at δ=26 (tab. 6).

δ	14	18	22	26	30	without
Error	2.24	2.01	2.02	1.98	2.03	2.32

Table 6: error rate in varing δ

In our method, the principal constants are the weight *W* which limits hand deformations and the threshold *G* defining the minimum value of a strong contour. The optimal values of *W* and *G* are defined with experimentations, with the second PCA set, according to error distance *e* and the number of miss detected hands. Its results are described in figure 7 and figure 9. The more the threshold increased, the more the detection is accurate but the more there are miss detections. A compromise must be adopted between *W* and *G* to have few miss detections and a great accuracy. The weight plays a very crucial role; it decreases in the same time, the error distance and the miss detection rate (fig. 9). Again one important value must be defined to complete the process: the number of points to describe a hand. With our method, nine points are constants (the five fingertip points and the four valley points between adjacent fingers). The numbers of points between its constants are fixed by experiments (fig. 8) with the optimal parameters δ, *W* and *G*. Of course, the error distance decreases when the computing time increases in adding points. So, the number of points is defined according to the system objective. Here, 12 inter points are used to define the model.

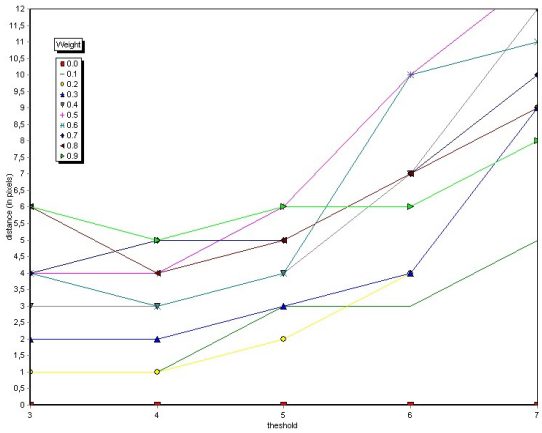

Figure 7: Miss detection rate

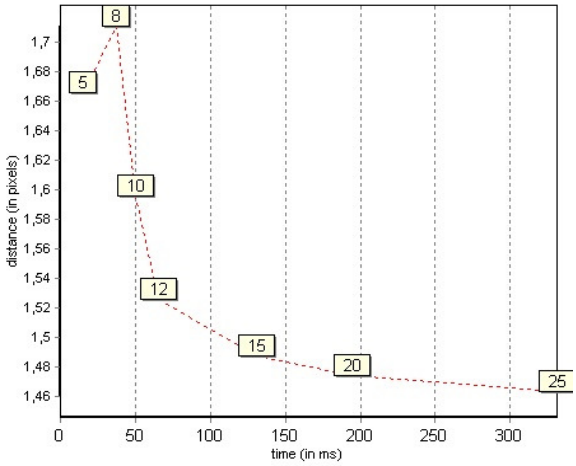

Figure 8: Hand accuracy and execution time according to the number of inter-points

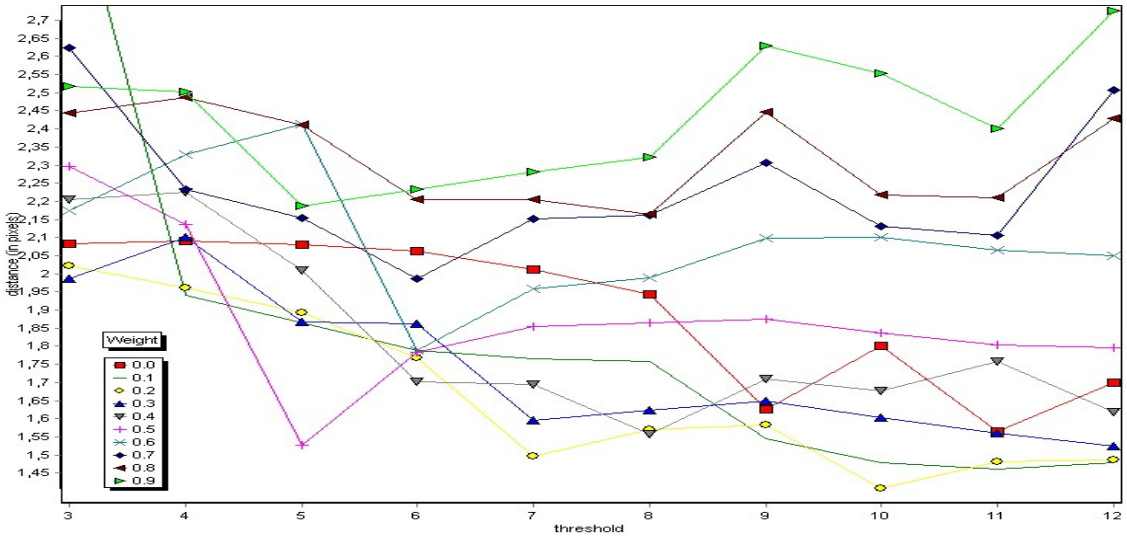

Figure 9: Error rate according to *W* and *G*

The robustness and the accuracy of the system is validated with the third ASM set which is a more difficult set than the second ASM set. The average error is 2.05 pixels with an identical miss detection rate, which confirm the parameters determination and the robustness of our system.

6. Conclusion

In this paper, a novel approach is presented to detect hand without tracking for contact less biometrics identification. The detection is based on a combination of skin detection and active shape model. The skin color is segmented by a multi-resolution neural network in RGB domain which initializes and constrains the active shape model. This model is deformed by an adjusted gradient determined in skin color space.

To assess the efficiency of our method, the Euclidian distance measures the difference between an ideal contour and the determined contour. An average error of 2 pixels on 454*341 images size with 4% of miss detected hands is obtained using our method.

To increase the speed and the robustness of the method, the hand detection process can be used with tracking. From the hand detection, some approaches based on texture and shape analysis [22] [23] have been used and show interesting results. Nevertheless, to obtain better results, a new recognition method must be defined for biometrics authentication.

7. Acknowledgement

This study is carried out in framework of LATEMS (Laboratoire de transactions électroniques, de monétique et de sécurité). The LATEMS is the common laboratory between ENSICAEN (Ecole Nationale Supérieure d'Ingénieurs de Caen) and FRANCE TELECOM R&D.

8. References

[1] Ming-Hsuan Yang; Kriegman, D.J.; Ahuja, N: "*Detecting faces in images: a survey*", Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume 24, Issue 1, 2002.

[2] Rowley, H.A.; Baluja, S.; Kanade, T: "*Neural-Network based face detection"*,CVPR'96, San Francisco, USA, 1996.

[3] Samaria, F.: "*face segmentation for identification using hidden markov models"*, Proceedings of 4th British Machine Vision, Surrey, UK,,1993.

[4] Jesorsky, O.; Kirchberg, K.J.; Frischholz R.: "*Robust Face Detection Using the Hausdorff Distance*", AVBPA 2001, Halmstad, Sweden, 2001.

[5] Garcia, C.; Delakis, M.: "*Convolutional Face Finder: A Neural Architecture for Fast and Robust Face Detection*", Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume 26, Issue 11, 2004.

[6] Kass, M.; Witkin, A.; Terzopoulos, D.: "*Snakes: Active contour models*", International Journal of Computer Vision, Volume 1, Issue 4, 1987.

[7] Cootes, T.F.; Taylor, C.J.: "*Statistical models of appearance for computer vision*", Technical report, University of Manchester, United Kingdom, 1999.

[8] Soriano, M.; Huovinen, S.; Martinkauppi, B.; Laaksonen, M.: "*Skin Detection in Video Under Changing Illumination Conditions*", Proc. 15th International Conference on Pattern Recognition, Barcelona, Spain, 2000.

[9] Jones, M.J.; Rehg, J.M.: "*Statistical Color Models with Application to Skin Detection*", International Journal of Computer Vision, Volume 46, Issue 1, 2002.

[10] Terrillon, J.-C.; Akamatsu, S.: "*Comparative Performance of Different Chrominance Spaces for Color Segmentation and Detection of Human Faces in Complex Scene Images*", 4th IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France, 2000.

[11] Anil K. Jain; Mao, J.; Mohiuddin, K. M.: "*Artificial neural networks: A tutorial*", IEEE Computer, Volume 29, Issue 3, 1996.

[12] Sural, S.; Qian, G.; Pramanik, S.: "*Segmentation and histogram generation using the HSV color space for image retrieval*", International Conference on Color Imaging X: Processing, Hardcopy, and Applications, San Jose, USA, 2005.

[13] Singh, S.K.; Chauhan, D.S.; Vatsa, M.; Singh, R.: "*A robust skin color based face detection algorithm*", International Tamkang journal of Science and Engineering, Volume 6, Issue 4, 2003.

[14] Brethes, L.; Menezes, P.; Lerasle, F.; Hayet, J.B.: "*Face tracking and hand gesture recognition for human-robot interaction*", ICRA'04, New Orleans, USA, 2004.

[15] Lui, T.Y.; Izquierdo, E.:"*Automatic Detection of Human Faces in Natural Scene Images by Use of Skin Color and Edge Distribution*", 5th International Workshop on Image Analysis for Multimedia Interactive Services, Lisbon, Portugal, 2004.

[16] Tomaz, F.; Candeias, T.; Shahbazkia, H.: "*Improved Automatic Skin Detection in Color Images*", International Conference on Digital Image Computing: Techniques and Applications, Sydney, Australia 2003.

[17] Di Zenzo, S.: "*A note on the gradient of a multi-image*", Computer Vision, Graphics and Image Processing, Volume 33, Issue 1, 1986.

[18] Smith, L. I.: "A tutorial on Principal Components Analysis", 2002.

[19] Han, C.-C.; Cheng, H.-L.; Lin, C.-L.; Fan, K.-C.: "*Personal authentication using palm-print features*", Pattern Recognition, Volume 36, 2003.

[20] Whitley, D.: "*A genetic algorithm tutorial*", Technical Report CS-93-103, Colorado State University, Department of Computer Science, 1993.

[21] Hsu, C.-W.; Chang, C.-C.; Lin, C.-J.: "*A Practical Guide to Support Vector Classification*", Technical Report, Department of Computer Science and information Engineering, National Taiwan University, 2003.

[22] Wai Kin Kong; David Zhang; Wenxin Li: "*Palmprint feature extraction using 2-D Gabor filters*", Pattern Recognition, Volume 36, Issue 10, 2003.

[23] Jain, K.; Ross, A.; Pankanti, S.: "*A prototype hand geometry-based verification system*", Conference on Audio and Video-based Biometric Person Authentication, Washington D.C., USA, 1999.

9. Glossary

ASM: Active Shape Model

DZA: Di Zenzo Algorithm

NN: Neural Network