Skin Detection using Neighborhood Information

Javier Ruiz-del-Solar and Rodrigo Verschae
Dept. of Electrical Engineering, Universidad de Chile, Santiago, CHILE
E-mail: jruizd@cec.uchile.cl, rverscha@ing.uchile.cl

Abstract

Skin detection is employed in tasks like face detection and tracking, naked people detection, hand detection and tracking, people retrieval in databases and Internet, etc. However, skin detection is not robust enough for dealing with some real-world conditions, like changing lighting conditions and complex background containing surfaces and objects with skin-like colors. This situation can be improved by incorporating context information in the skin detection process. For this reason in this article a skin detection approach that uses neighborhood information is proposed. A pixel will belong to the skin class only if a direct neighbor does. This idea is implemented through a diffusion process. Two new algorithms implementing these ideas are described and compared against state-of-the-art skin detection algorithms.

1. Introduction

Skin detection or segmentation is a very popular and useful technique for detecting and tracking human-body parts, specially faces and hands. Its most attractive properties are: (i) high processing speed due to its low-level processing, and (ii) invariance against rotations, partial occlusions and changes in pose. However, skin detection is not robust enough for dealing with complex environments. Changing lighting conditions, and complex background containing surfaces and objects with skin-like colors are major problems, limiting its use in practical real-world applications. For solving the mentioned drawbacks many groups have centered their research on selecting the color-space most suitable for skin detection. Many different color models have been employed, among them: RGB [14][20], normalized RGB [3][16], HIS-HSV [10][15], YCbCr [7][4], YIQ [6], YES [13], YUV [1], CIE XYZ [5], CIE LUV [20], and ab [10].

Other authors have centered their efforts on modeling the physical properties of the acquisition process. In [12] the physical properties and the calibration of the camera to be employed in the acquisition process are modeled, while in [7] a color constancy method (white balancing) for obtaining some invariance against changing illumination is employed. We believe that these approaches are not general enough and they can be implemented only in some special cases.

Some other authors have used statistical models for solving the skin/non-skin classification problem. Most successful approaches are Mixture of Gaussians (MoG) [20] and histogram models [9]. They differ in the parametric or non-parametric form of computing the skin/non-skin probabilities. (Although is a popular belief that histogram models have a high processing speed because they employ LUTs (look-up-tables), the same procedure can be used for the MoG by storing the evaluated Gaussian functions in the LUTs). We think that statistical models are in the right direction for dealing with the real-world problems of skin detection, however they miss the benefits of using contextual information.

All mentioned approaches are based on the same *pixel-wise* processing paradigm, in which each image pixel is individually analyzed. We think this paradigm should be extended; context information should be incorporated in the skin detection process. Human beings can detect skin in real scenes, or in pictures and videos without problems. However, for a human being the classification of a single pixel as skin or non-skin is a very difficult task. We believe the reason is that human skin detection is not a simple low-level process but a process in which high-level mechanisms are also involved. If we think in the human perception of a blue ball, we will agree in that the ball is perceived blue as a whole, and not as a ball having blue patches and some other color patches produced by differences on illumination. For having this kind of perception not only low-level color processing mechanisms for blue pixels and patches detection are involved, but also shape detection mechanisms for detecting the ball and mechanisms for color constancy and interpolation [18]. In the same way, the detection of skin in a face or in a hand does not involve only low-level color processing mechanisms, but it also incorporates high-level processes to assist the detection of skin (detection of hair, detection of clothes, etc), and also some spatial diffusion mechanisms, employed in any human segmentation process of colors and textures [18]. Taking this fact into consideration we propose the use of context information in the skin detection process. Spatial context can be employed for skin detection in still images, while time context can also be employed when detecting skin in videos. When dealing with high-level tasks such as face and person detection, high-level information (e.g. detected faces) can assist the skin detection process.

Following this idea in this article is proposed a skin segmentation approach that uses neighborhood information, i.e. the decision about the class (skin or non-skin) of a given pixel considers information about the pixel's neighbors. Although this idea has been employed before in the post-processing of skin-detections (e.g. holes in skin detected areas are filled using morphological dilation and isolated skin pixels are deleted using morphological erosion) and more recently in [2] using unsupervised segmentation and region grouping, in this work a different and more radical approach is employed: a diffusion process is implemented for determining the skin pixels. The aim of this process is not just the grouping of neighbor skin pixels, but also the determination of skin areas where the "skiness" of pixels are larger than a minimal threshold and where the "skiness" of neighbor pixels changes smoothly. This process is controlled using three thresholds: one for determining the minimal acceptable "skiness" of a skin pixel, one for controlling smooth changes of the pixel "skiness" between neighbors, and a third one for determining the seeds of the diffusion process. Before diffusion the "skiness" measure, which can either correspond to the probability of being a skin pixel when a statistical model is used, or to a membership degree when a fuzzy approach is employed, should be calculated. Even though a diffusion process is employed for implementing the skin segmentation, a reasonable high processing speed is achieved.

The article is structured as follows. The proposed skin detection approach is described in section 2. In section 3 comparisons with other algorithms using real-world images are presented. Finally, in section 4 some conclusions of this work are given.

2. The proposed Skin Detection Approach

As mentioned in the introduction, the here proposed fuzzy skin detection approach has two steps: (i) Pixel-Wise Classification, and (ii) Controlled Diffusion. This idea can be implemented using different diffusion strategies, diverse statistical and fuzzy models, as well as different color spaces. For this reason several algorithms can be derived. In this section some of them will be described.

For the sake of simplicity in the description of our algorithms we will use the spatial-range domain as image space. In this domain each image pixel has two parts, a spatial and a range part, where the range part may be written as a function of the spatial part:

$$x_j = (x_j^s, x_j^r) = (x_j^s, I(x_j^s)) \qquad (1)$$

The superscripts *s* and *r* denote the spatial and range parts of the pixels. The spatial range domain has a dimension of *d*=*r*+2, with *r*=1 for grayscale images and *r*=3 for color images. The value "2" represents the bi-dimensional spatial characteristic of images.

2.1. Pixel-Wise Classification

As mentioned the pixel-wise classification can be either probabilistic or fuzzy. Thus the belonging of a pixel to the skin class is calculated as a real number between 0 and 1. When a statistical model is used, this number corresponds to the probability of being a skin pixel. When a fuzzy approach is employed, it corresponds to a fuzzy membership degree. In this last case a given crisp pixel-wise classification algorithm can be employed as a basis algorithm, and their results fuzzificated. For instance, if an elliptical skin cluster is used as pixel-wise classification method (on a given color space), before classification the cluster is fuzzificated: a value 1 is assigned to the center of the cluster, and decreasing values, which depend on their distances to the center, are assigned to the rest of points. When a probabilistic model is chosen, the best alternatives are MoG and histogram models, being their main difference the way in which the skin/non-skin probabilities are calculated. MoG is a parametric model and thereafter less training samples are required for obtaining good probabilities' estimators. Usually it is argued that histograms are faster because they are implemented using LUTs [9], but LUTs can also be used for implementing the MoG. Although in [9] slightly better results are obtained using histograms than using MoG, the reason seems to be the use of a non-skin model in the former case.

A general pixel-wise classification algorithm is presented in figure 1. As it can be seen, this algorithm imposes no conditions on the color space, or on the statistical model or fuzzification function being used. We implemented the following pixel-wise skin classification algorithms:

2.1.1. RGB-MoG. We employ the RGB color space and a mixture density function. In particular we use the function proposed in [9]:

$$g_{skin}(x) = \sum_{j=1}^{16} w_j \frac{1}{(2\pi)^{3/2} \left|\Sigma_j\right|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu_j)^T \Sigma_j^{-1} (\mathbf{x}-\mu_j)} \qquad (3)$$

where the contribution of the jth Gaussian is determined by a scalar weight w_j, mean vector μ_j and diagonal covariance matrix Σ_j.

2.1.2. Fuzzy-YCbCr. We employ the YCbCr color space and we first define an elliptical, crisp, skin cluster in the CbCr plane. The exact parameters of the elliptical model can be obtained by training. Second, we fuzzyficate the cluster: a value 1 is assigned to the center of the cluster and a value 0.5 is assigned to the border. Intermediate values are given to other areas of the cluster using a lineal model, which assigns decreasing values from the cluster center to the cluster border. Outside the cluster a second lineal model, with other slope is employed. Non-lineal models can also been used.

```
pixel_wise_classification(I, I_skin) {
        foreach $x_i \in I, y_i \in I_{skin}$
                $y_i^r = g_{skin}(x_i^r)$
}
```

I: input image; I_{skin}: image with skin probabilities or membership degrees as pixel values.
g_{skin}: probabilistic model or fuzzyfication function.

Figure 1. Pixel-wise classification algorithm.

2.2. Controlled Diffusion

The final decision about the pixel's class is taken using a spatial diffusion process that takes into account context information. In this process a given pixel will belong to the skin class if and only if its Euclidean distance, calculated in a given space, with a direct diffusion-neighbor that already belongs to the skin class, is smaller than a certain threshold (T_{diff}). The seeds of the diffusion process are pixels with a high probability of being skin or with a large membership to the skin class, i.e. their probability or membership degree is larger than a given threshold (T_{seed}). The extension of the diffusion process is controlled using a third threshold (T_{min}), which defines the minimal probability or membership degree allowed for a skin pixel. Two different algorithms were implemented, where the similarity between neighbor pixels is measured in different spaces (see figure 2):

2.2.1. Diffusion1. The similarity between neighbor pixels (8-connectivity) is measured as the Euclidean distance in a certain color space.

2.2.2. Diffusion2. The similarity between neighbor pixels (8-connectivity) is measured as the absolute value of the difference of the degree of membership to the skin class of each pixel.

```
diffusion_algorithm(I_skin, S_skin, I){
        find_seeds(I_skin, S_seed)
        foreach $s \in S_{seed}$
                PushItem(S_skin, s)
                diffusion_rec(s, S_skin, I)
}
find_seeds(I_skin, S_seed) {
        foreach $x_i \in I_{skin}$
                if $x_i^r > T_{seed}$
                        PushItem(S_seed, x_i)
}
diffusion_rec(s, S_skin, I){
        foreach $x_j^s \in ConnectedNeighbors\ (s^s)$
                if $x_j \notin S_{skin}$ //Choose either Diff1 or Diff2
                        if $|I(x_j^s) - I(s^s)| < T_{diff} \wedge x_j^r \geq T_{\min}$ //Diff1
                        /* if $|x_j^r - s^r| < T_{diff} \wedge x_j^r \geq T_{\min}$ //Diff2*/
                                PushItem(S_skin, x_j)
                                diffusion_rec(x_j, S_skin)
}
```

I_{skin}: image with skin probabilities or membership degrees as pixel values. S_{skin}: final set of skin pixels, the output of the algorithm. S_{seed}: set of seed pixels. *I*: input image.
PushItem(*S*,*x*) adds *x* to the end of *S*.
ConnectedNeighbors(*s*) returns the 8-connected neighbors of *s* in the 2D lattice of the image.

Figure 2. Diffusion algorithm.

3. Results

For testing the performance of the proposed algorithms, in real-world conditions, we selected a set of 27 images (altogether 5,939,422 pixels) obtained from Internet and from digitized news videos from ABC and CNN. The selected images are considered difficult to segment (skin/not skin), because they have either changing lighting conditions or complex backgrounds containing surfaces and objects with skin-like colors are major problems. The selected images, as well as, their ground truth information (labeled skin pixels) are available for future studies in [22].

The algorithms to be compared are:

- Diff1RGBMoG-RGB, which uses the *RGB-MoG* pixel-wise classification algorithm (see 2.1.1) and the *Diffusion1* algorithm (see 2.2.1).

- Diff2RGBMoG-MoG, which uses the *RGB-MoG* pixel-wise classification algorithm (see 2.1.1) and the *Diffusion2* algorithm (see 2.2.2).

- Jones1, which corresponds to the mixture of Gaussians classifier proposed in [9], using only the skin color model and a fixed output threshold T_{fix1}.

- Jones2, which corresponds to the mixture of Gaussians classifier proposed in [9], using the skin color and the non-skin color models and a fixed ratio threshold T_{fix2}.

- Hsu1, which corresponds to the skin detection algorithm proposed in [7] (YCbCr, elliptical cluster model), but without use of whitening compensation, for fairness in the comparison with the other algorithms.

In table 1, the ranges of the different parameters used to obtain the ROC curves and operation points displayed in figure 3 and 4, are shown. For obtaining the ROC curves for the Diff1RGBMoG-RGB algorithm the following procedure was follow: for a given value of T_{min} and T_{diff}, T_{seed} was moved from 0.15 to 0.55. For the Jones1 and Jones2 algorithms the parameters T_{fix} and T_{fix2} were modified. For the Hsu1 algorithm the radios that define the ellipse were modified proportionally, obtaining concentric ellipses that maintain the ratio of the radios and the orientation of the original ellipse.

In figure 3.a different operation points of the Diff1RGBMoG-RGB and Diff2RGBMoG-MoG algorithms are shown. Both algorithms have similar performance, and each of them gives better results than the other in a different range of false positives, although Diff1RGBMoG-RGB has a better performance for a detection rate greater than 50%. For this reason this algorithm was used for the comparisons with Jones1, Jones2 and Hsu1. In figure 3.b the ROC curves of Diff1RGBMoG-RGB are shown. In this figure, it can be noticed the effects of the parameters (sensibility) in the performance of the algorithm. A too low value of T_{min} makes the algorithm to connect skin areas with non-skin areas, independently of the values of T_{diff} and T_{seed}. T_{seed} variations change almost linearly the performance of the algorithm. If T_{diff} is lower than 20, the performance of the algorithm is almost no sensible to changes in this parameter.

Finally, in figure 4 are shown the ROC curves of the compared algorithms. It can be noticed that the here proposed algorithms outperform the results of previous works. For instance, for a detection rate of 70%, Diff1RGBMoG-RGB has a half of false positive rate of Hsu1 and _ of the false positive of Jones1.

Table 1. Algorithms and parameters.

Algorithms	**Parameters**	**Values**
Diff1RGBMoG-RGB	MoG parameters	Obtained from [9]
	T_{seed};T_{diff};T_{min}	$T_{seed} \in [0.1, 0.8]$ $T_{min} \in [0.04, 0.40]$ $T_{diff} \in [10, 20]$
Diff2RGBMoG-MoG	MoG parameters	Obtained from [9]
	T_{seed};T_{diff};T_{min}	$T_{seed} \in [0.1, 0.7]$ $T_{min} \in [0.04, 0.25]$ $T_{diff} \in [16, 30]$
Jones1	MoG parameters	Obtained from [9]
	T_{fix1}	T_{fix1} [0.2, 1]
Jones2	MoG parameters	Obtained from [9]
	T_{fix2}	T_{fix2} [0.02, 2]
Hsu1	Elliptical model	Obtained from [7]

4. Conclusions

In this article a skin detection approach that uses neighborhood information was proposed. Under this approach a pixel belongs to the skin class only if it has a probability of belonging to the skin class over a certain threshold, and if some of its neighbors, previously classified as belonging to the skin class, is similar to it. This classification is implemented through a spatial diffusion process started from pixels with high skin probability (seeds). Two algorithms implementing these ideas were described and favorable compared with state of the art skin detection algorithms using real-world images.

It is worth to mention that the here proposed approach for skin detection can be employed with any pixel-wise skin detection algorithm. It is only necessary to obtain a "skiness" value for each image pixel, for instance by fuzzifying a crisp algorithm, and then to apply the here proposed spatial diffusion algorithm.

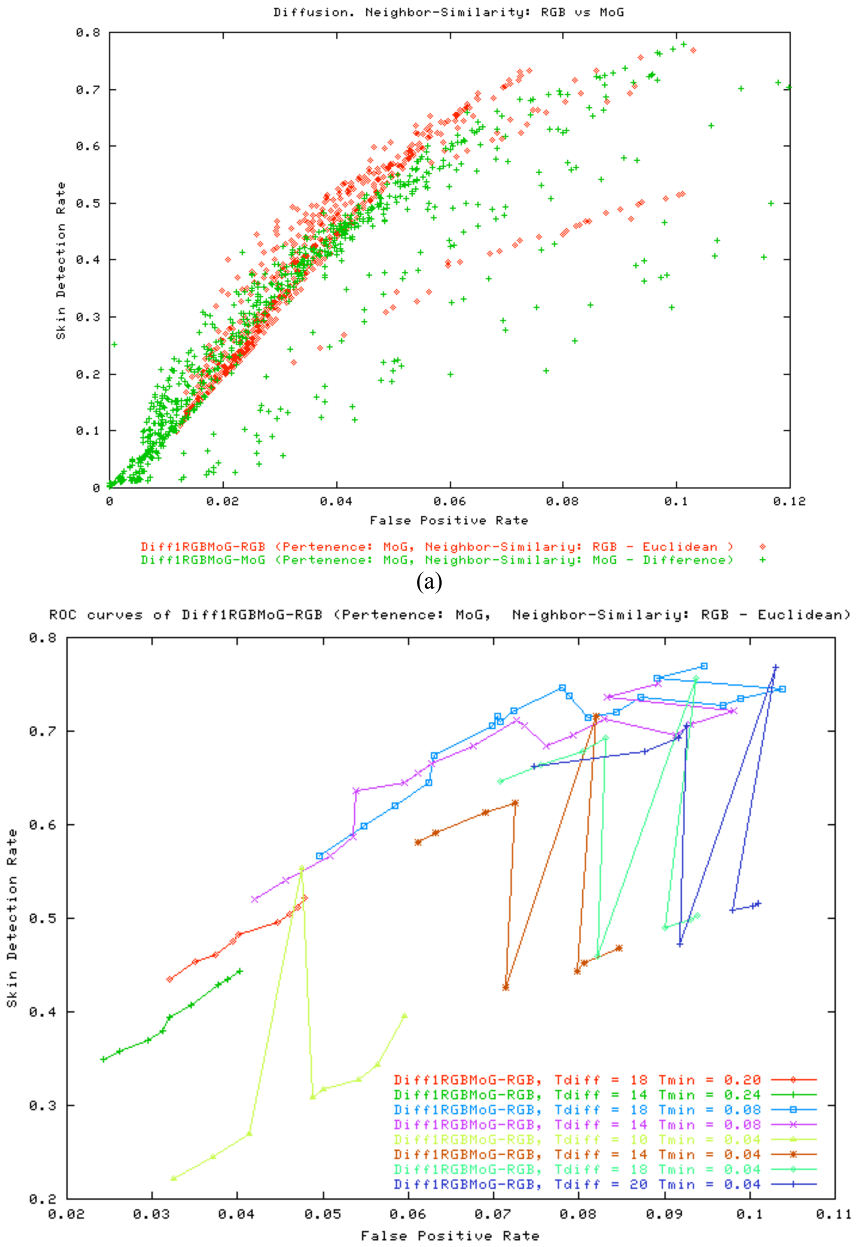
Acknowledgement

This research was funded by the FONDECYT Project 1030500 (CONICYT, Chile).

5. References

[1] M. Abdel-Mottaleb, and A. Elgammal, "Face detection in complex environments from color images", *Proc. IEEE Int. Conf. on Image Processing*, Kobe, Japan, 3: 622-626, 1999.

[2] A. Albiol, L. Torres, Ed. Delp, "An Unsupervised Color Image Segmentation Algorithm for Face Detection Applications", *IEEE Int. Conf. on Image Proc. – ICIP 2001*, Greece, 2001.



(a)

(b)

Figure 3. (a) Operation points of Diff1RGBMoG-RGB and Diff1RGBMoG-MoG. (b) Some ROC Curves Diff1RGBMoG-RGB different curves.

[3] L.M. Bergasa, M. Mazo, A. Gardel, M.A. Sotelo and L. Boquete, "Unsupervised and adaptive Gaussian skin-color model", *Image and Vision Computing* 18(12): 987-1003, 2000.

[4] D. Chai, and K.N. Ngan, "Locating facial region of a head-and-shoulders color image" *Proc. 3rd IEEE Int. Conf. on Automatic Face and Gesture Recognition*, Nara, Japan, 124-129, 1998.

[5] Q. Chen, H. Wu, and M. Yachida, "Face detection by fuzzy pattern matching", *Proc. 5th Int. Conf. on Computer Vision*, Cambridge, Massachusetts, USA, 591-596, 1995.

[6] Y. Dai, and Y. Nakano, "Extraction for facial images from complex background using color information and SGLD matrices", *Proc. 1st Int. Workshop on Automatic Face and Gesture Recognition*, Zurich, Switzerland, 238-242, 1995.

[7] R. L. Hsu, M. Abdel-Mottaleb, and A.K. Jain, "Face detection in color images", *IEEE Trans. on Pattern Anal. and Machine Intell.* 24(5): 696-706, 2002.

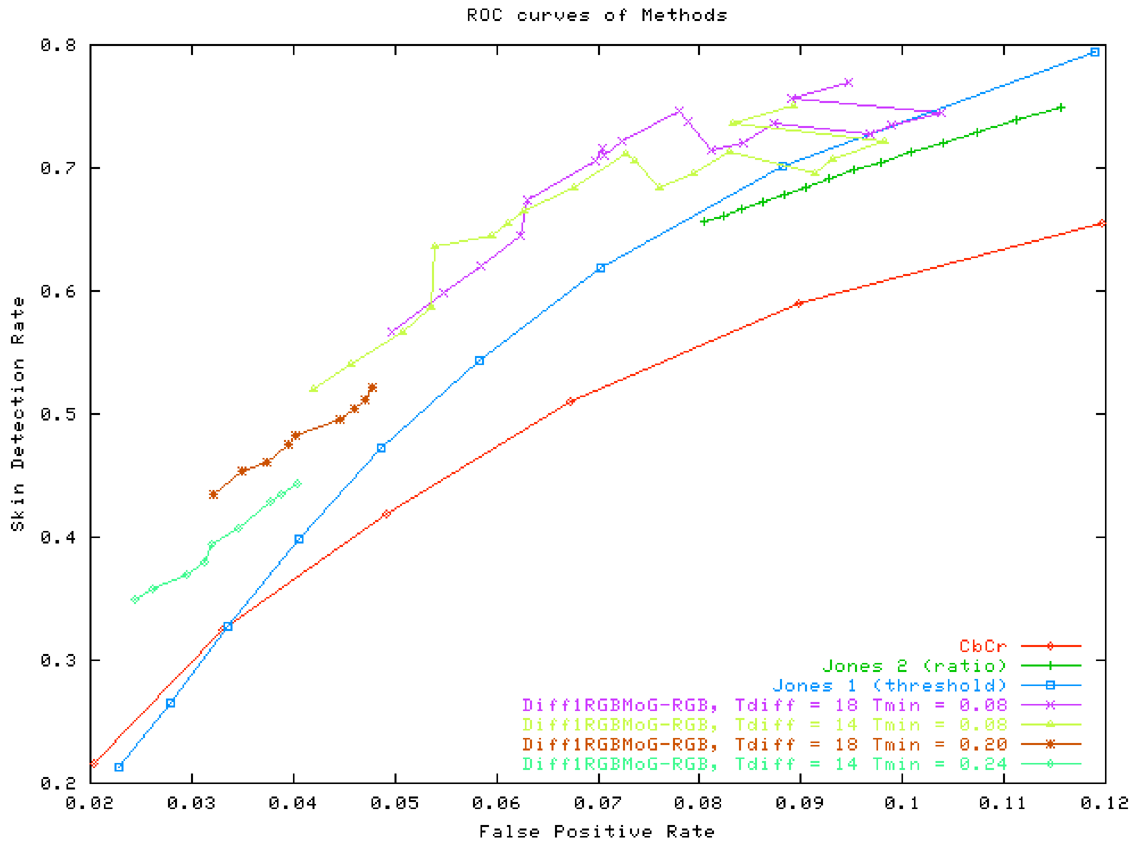


Figure 4. ROC Curves of the compared algorithms.

[8] T. S. Jebara, and A. Pentland, "Parameterized structure from motion for 3d adaptive feedback tracking of faces", *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 144–150, San Juan, Puerto Rico, June 17-19, 1997.

[9] M.J. Jones, and J.M. Rehg, "Statistical color models with application to skin detection", *Int. Journal of Computer Vision* 46(1): 81-96, 2002.

[10] S. Kawato, and J. Ohya, "Real-time detection of nodding and head-shaking by directly detecting and tracking the between eyes", *Proc. 4th IEEE Int. Conf. on Automatic Face and Gesture Recognition*, Grenoble, France, 40-45, 2000.

[11] R. Kjeldsen, and J. Kender, "Finding skin in color images", *Proc. 2nd International Conference on Automatic Face and Gesture Recognition*, Killington, Vermont, USA, 312-317, 1996.

[12] B. Martinkauppi, *Face Color under Varying Illumination – Analysis and Applications*, Doctoral Thesis, university of Oulu, Finland, 2002.

[13] E. Saber, and A.M. Tekalp, "Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions", *Pattern Recognition Letters* 17(8): 669-680, 1998.

[14] S. Satoh, Y. Nakamura, and T. Kanade, "Name-it: naming and detecting faces in news videos", IEEE Multimedia 6(1): 22-35, 1999.

[15] D. Saxe, and R. Foulds, "Toward robust skin identification in video images", *Proc. 2nd Int. Conf. on Automatic Face and Gesture Recognition*, Killington, Vermont, USA, 379-384, 1996.

[16] Q. B. Sun, W.M. Huang, and J. K. Wu, "Face detection based on color and local symmetry information", *Proc. 3rd IEEE Int. Conf. on Automatic Face and Gesture Recognition*, Nara, Japan, 130-135, 1998.

[17] J. Ruiz-del-Solar, A. Shats, and R. Verschae, "Real-Time Tracking of Multiple Persons", *Proc. of the 12th Int. Conf. on Image Analysis and Processing – ICIAP 2003*, September, Mantova, Italy, 2003.

[18] L. Spillman and J. Werner (Eds.), *Visual Perception: The Neurophysiological Foundations*, Academic Press, 1990.

[19] J. Yang, W. Lu, and A. Waibel, "Skin-color modeling and adaptation", *Proc. 3rd Asian Conference on Computer Vision*, 687–694, Hong Kong, China, January 1998.

[20] M. H. Yang, and N. Ahuja, "Detecting human faces in color images", *Proc. IEEE Int. Conf. on Image Processing*, Chicago, Illinois, USA, 1: 127-130, 1998.

[21] T.W. Yoo, and I. S. Oh, "A fast algorithm for tracking human faces based on chromatic histograms", Pattern Recognition Letters 20(10): 967-978, 1999.

[22] http://skin.li2.uchile.cl/db1