

# Data-Mining Based Skin-Color Modeling Using the ECL Skin-Color Images Database

Mohamed Hammami, Dzmitry Tsishkou, and Liming Chen

LIRIS, FRE 2672 CNRS, Ecole Centrale de Lyon  
36, Avenue Guy de Collongue  
69131 Ecully, France

{mohamed.hammami, dzmitry.tsishkou, liming.chen}@ec-lyon.fr

**Abstract.** Many human image processing techniques use skin detection as a first stage in subsequent feature extraction. In this paper we describe methods of skin detection using a data-mining technique. We also show the importance of the choice of the base simple to the performance of our skin analysis techniques. We present the details and the process of construction of our database which we have called “the ECL Skin-color Images Database from video”. We will show that the use of a database derived from live video gives better results than one derived from internet images for face detection in video application.

## 1 Introduction

Skin detection can be defined as the process of selecting which pixels of a given image correspond to human skin. Skin is arguably the most widely used primitive in human image processing research, with applications ranging from face detection[10] and person tracking to pornography filtering[5]. Skin detection techniques can be both simple and accurate, and so can be found in many commercial applications, for example the driver eye tracker developed by Ford UK [9].

Most potential applications of skin-color model require robustness to detect significant variations in races, differing lighting conditions, textures and other factors. This means large databases composed of tens of millions of pixels are necessary in the training stage in order to construct an effective skin color-model. It is a major challenge to manage such a complex structure. One of the driving applications for skin-color model construction of large datasets is data-mining.

Construction of a skin-color database may be done in a different way, depending on an application. Perhaps the four most significant factors are: (1) the size of the database; i.e. it should contain more than 10 million of skin-color pixels. (2) the variety in the image content, so that it should be representative for changes in race, sex, and environmental conditions. (3) the database source, that is particular depends on the application, such as www adult content filtering, face detection for security systems or face detection for video. Although there is a comprehensive database of skin-color images with a large number of subjects, and with significant race, sex and environmental conditions variations (Jones et al., 1998).

In this paper we propose a data-mining method for skin-color modeling. Our system uses the Ecl Skin color data base and SIPINA method [13] to automatically extract the most efficient parameters of the skin-color model.

The remainder of this paper is organized as follows. The Construction of the ECL skin color image database is presented in Section 2. The data-mining based skin-color modeling is presented in Section 3. The face detection system is presented in Section 4. Experimental evaluation of the proposed skin-color model is discussed in section 5. Section 6 draws conclusions on our work.

## 2 Data Collection: The ECL Skin-Color Database

Between April 2002 and June 2002 we collected a database of over 1110 skin-color images of more than 1110 people. (the total size of database is about 91 MB). We call this database the ECL Skin-color Images from Video (SCIV) database (Fig.1). To obtain wide variations in race, sex and environmental conditions we proceed more than 30 hours of TV news (Euronews, TF1, France2). For capturing Euronews TV stream we used Pinnacle TVSat hardware, for TF1 and France2 internal TV receiver was used together with Hauppauge WinTV USB video tuner. All video were stored in MPEG4 compatible format using 320x240 resolution on image size, no audio, and medium compression settings. To obtain binary masks we used skin-color segmentation tool similar to one constructed by (Jones et al.,1998). We build a similar histogram based skin-color model and implement a software for skin-color images segmentation. Finally we manually corrected binary masks in Adobe Photoshop 7 software. Besides the two major partitions of the database (skin-color images and corresponding binary masks). The ECL SCIV database therefore consists of two major partitions, the first skin-color images, the second corresponding binary masks.

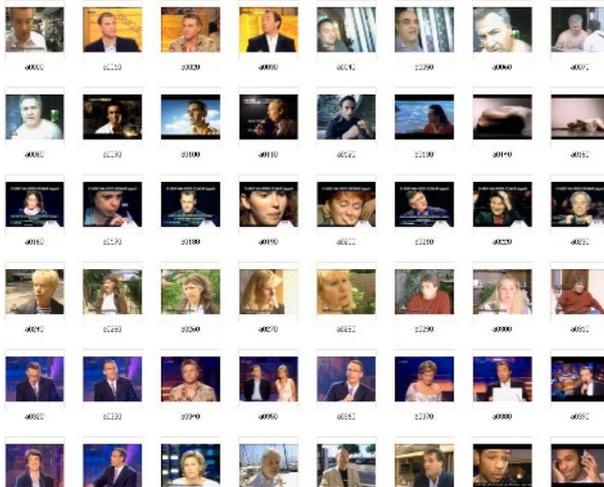


Fig. 1. Sample skin-color images from ECL SCIV database

**2.1 Capture Apparatus**

Obtaining images of skin-color from video requires special capturing hardware. Image quality depends on the video image type. There are two sources for ECL SCIV database: (1) video images captured from a satellite using Pinnacle TVSat hardware, which enables to receive and store MPEG2 stream on the hard disk, and (2) video images captured from an internal receiver with Hauppauge WinTV USB hardware. The first solution represents digital TV standards, using PCI bus transfer it makes possible to store high quality video images. Meanwhile USB based tuner accommodates analogue TV so that medium image quality within low bandwidth transfer. Finally all video, with the total duration of 30 hours, was converted to MPEG4 compatible format using DivX 5.1 codec with no audio option.

**2.2 Capture Procedure**

To obtain significant variation in face expressions, pose, position, race, sex and environmental we did manual shot segmentation and extracted one frame per each shot, containing skin-color with total area size more than 5% of the whole image size. So that we extracted 1100 skin-color images from 30 hours of the video. The primary subject of our research activity is the skin detection in video (particular TV news) application; therefore we captured 15 hours of Euronews with a high quality. The TF1 and France2 were captured during of 15 hours, covering the whole range of TV programs types, including TV news, sitcoms, movies, sport (World cup 2002), commercials and other.

**2.3 Skin-Color Segmentation**

Each skin-color image in ECL SCIV was proceed in the following manner: first skin-color regions were manually labelled using software tool shown on figure 2. This tool allows user to interactively segment skin-color regions by controlling the threshold of histogram based skin-color model pixel classifier.

The threshold slider controls the accuracy of segmenting the region of interest. In labelling we were careful to exclude the eyes, hair and mouth opening.



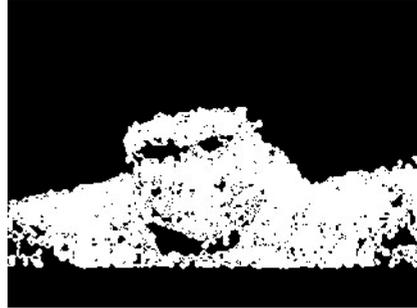
**Fig. 2.** Skin-Color segmentation tool

Second each binary mask was manually corrected in Adobe Photoshop 7 image editor.

Finally all segmented images also called as masks were stored in independent files with name corresponding to its origin images with white pixels representing skin-color pixels and black pixels –non skin color (Fig.4). Hence we are able to extract only skin color or non skin color pixels from image using mask automatically.



**Fig. 3.** Skin-color image from ECL SCIV database



**Fig. 4.** Binary mask, corresponding to the skin-color image on Fig.3

## 2.4 Database Organization

On average the capture, preprocessing and indexing took about 10 minutes per each image. The images are 320x240 color images 24BPP. Each image is compressed using the JPEG codec with high quality. The database is organized in 3 partitions, the first consisting of the skin-color images (manually extracted from video), the second consisting the corresponding binary masks (automatically segmented and manually corrected from corresponding skin-color images) and the third one – meta-data index (manually done skin-color images classification).

## 2.5 Meta-Data

Regarding potential uses of the database, besides the two major portions of the database, we also collected a set of meta-data to aid further ECL SCIV database analysis:

**Race:** We classify each image into European, Asian, Africa, Latin and Arabic. The race distribution graph shows that the database consists of 82% Europeans, 8% Africans, 5% Latin, 3% Asian and 2% Arabic people (Fig 5).

**Gender:** We manually did men/woman classification on ECL SCIV database. The results demonstrate 72% of men and 28% of women populating the database (Fig 6).

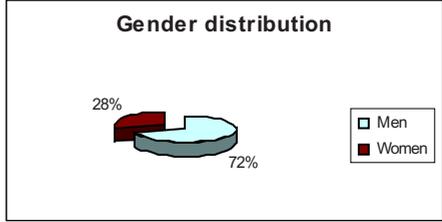
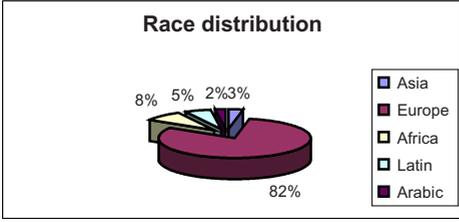


Fig. 5. ECL SCIV database: race distribution      Fig. 6. ECL database: Gender distribution

### 3 Data-Mining Based Skin-Color Modeling

#### 3.1 Data Preparation

Large databases composed of tens of millions of pixels are necessary to explore various types of lighting conditions, races, etc. in order to construct an effective skin color-model. Several color spaces have been proposed in the literature for skin detection applications. YCbCr has been widely used since the skin pixels form a compact cluster in the Cb-Cr plane. As YCbCr is also used in video coding and then no transcoding is needed, this color space has been used in skin detection applications where the video sequence is compressed [1, 10]. In [11] two components of the normalized RGB color space (rg) have been proposed to minimize luminance dependence. And finally CIE Lu\*v\* has been used in [12]. However, it is still not clear which is the color space where the skin detection performance is the best.

To create skin-color model using color information we use ECL SCIV database; all intensity information on color pixels was extracted using binary masks.

For each pixel we compute its representation in following normalized colorspaces: RGB, HSV, YIQ, YCbCr, CMY in order to find the most discriminative set of color axes. To store the dataset we include information on pixel values in different color-spaces with corresponding binary masks values indicating the presence of either skin-color or non-skin color pixels[4].

#### 3.2 Skin-Color Modeling: SIPINA

A number of classification techniques from the statistics and machine learning communities have been proposed [3, 7, 8]. A well-accepted method of classification is the induction of decision trees [2, 7]. A decision tree is a flow-chart-like structure consisting of internal nodes, leaf nodes, and branches. Each internal node represents a decision, or test, on a data attribute, and each outgoing branch corresponds to a possible outcome of the test. Each leaf node represents a class. In order to classify an unlabeled data sample, the classifier tests the attribute values of the sample against the decision tree. A path is traced from the root to a leaf node which holds the class

predication for that sample. Decision trees can easily be converted into IF THEN rules and used for decision-making.

SIPINA [13] is a widely used technique for data-mining. The effectiveness of SIPINA is superior to classical methods such as ID3 and C4.5[8], because the distribution equivalency can be considered in a population-wise manner, irrespective of any fixed solutions proposed previously. This accounting mechanism accurately charts usage distribution and leads to the highest performance among other methods, particularly for skin-color modelling. Let the set of pixels be extracted and pre-processed automatically from training images and corresponding binary masks. Each pixel  $W$  is associated with its  $C(w)$  class (skin-color, non skin-color).

$$C : \Omega \rightarrow \mathcal{C} = \{\text{skin-color, non skin-color}\}. \quad (1)$$

$$W \rightarrow C(w). \quad (2)$$

The observation of  $C(w)$  is not easy, because of lighting conditions, race differences and other factors. All these factors complicate the skin-color classification process. Therefore we are looking for mean value to describe class  $C$  of each element in different color spaces.

As a result of applying this method to a training set, a hierarchical structure of classifying rules of the type "IF...THEN..." is created.

Using the ECL data base of skin-color and non-skin color images and a data-mining technique we discover that HSV is the most discriminative colorspace. Detailed values on skin-color model decision rules and experimental results on the training process will be published in our extended technical report, however we will present experimental results on evaluation of this model for face detection in video.

## 4 Face Detection in Video Analysis and Indexing

Rapid growth of telecommunication data is caused by tremendous progress in information technology development and its applications. Modern figures of the volume of video data are ranged from hundred hours of home video for personal use to a million-hour TV company's archive. The condition of a successful practical use of this data is the stringent enforcement in the following areas: storage, transmission, preprocessing, analysis, indexing. Despite the fact that information storage and transmission systems are capable to supply video data to user there is still a developmental gap between the storage and the indexing. A problem of manual video data preprocessing, analysis and indexing has no practical solution in terms of human perception and physiological abilities. Therefore the construction of the fully automatic video indexing system is the current research subject. High level semantic analysis and indexing systems generally exploit some known specific properties of the video, e.g., that the video is a hierarchical multi-layer combination of background/foreground objects. One can decompose the video into the set of such objects and the structure. That requires video structure understanding and object recognition solutions. Examples of such object recognition solutions include face

detection and recognition among others. There are several reasons why one is interested in a use of facial information for video analysis and indexing. First of all, visual facial information serves for clue up on a personal identity. That is one of the most informative tags for a particular video invent. The second reason why the face detection and recognition is important is that it has been given a great attention from the scientific society until present. It’s a strong advantage, taking into account the fact that the number of potential objects to be recognized by the video analysis and indexing system is nearly infinite. Finally, the most important reason why facial information is important is that it enables us to dramatically increase the number of applications of the video analysis and indexing.

## 5 Experimental Evaluation

In this section we will deal with practical implementation of the skin-color model for face detection application. We performed two experiments and present the results here. All experiments were performed with video captured to files, so that pure performance (faster than real time) could be evaluated. The skin-color models used in experiments were: data-mining based using the ECL SCIV derived from live video (A), data-mining based using the data base derived from internet images [6] (B). We note that the face detection in video system was tuned for fast performance and some features were disabled. Meanwhile skin-color preprocessing takes less than 1% of total computational complexity; therefore the only parameter evaluated is a total number of detected persons per hour of video (table 1).

**Table 1.** Face detection in video system: parameters evaluation

Experimental conditions:	Full time French public TV (TF1, France2, France3, M6)
Duration	30 hours
Time interval	15 days
Skin-color pixels (per hour)	A-24770831 B-21615935
Skin-color regions (per hour)	A-1228209 B-1050561
Correct face images (detected per hour)	A – 2169 B – 2021
Performance	Average fps – 78 (Intel Pentium 2.0 GH)
Minimum size	20x20 pixels
Maximum size	infinite
In plane Rotation freedom	Up to 25 degrees

## 6 Conclusions

We have introduced in this paper a new skin-color modeling approach – data-mining based skin-color model. We presented the details and the process of construction of the “ECL Skin-color Images from Video Database”, which was used a training dataset. Experimental evaluation has shown that data mining-based skin-color model could be integrated into the face detection in video system, providing acceptable detection results. Usage of a special training dataset slightly improves total performance of face detection in video system. Further research activities would be concentrated on a real-time data mining-based skin-color model adaptation to the lighting conditions.

## References

1. A. Albiol, L. Torres, C.A. Bouman, and E. J. Delp, “A simple and efficient face detection algorithm for video database applications,” in *Proceedings of the IEEE International Conference on Image Processing*, Vancouver, Canada, September 2000, vol. 2, pp. 239–242.
2. L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification of Regression Trees*. Wadsworth, 1984.
3. U. M. Fayyad, S. G. Djorgovski, and N. Weir. Automating the analysis and cataloging of sky surveys. In U. Fayyad, G. PiatetskyShapiro, P. Smyth, and R. Uthurusamy, editors, *Advances in Knowledge Discovery and Data Mining*, pages 471–493. AAAI/MIT Press, 1996.
4. M. Hammami, Y. Chahir, L. Chen, D. Zighed, "Détection des régions de couleur de peau dans l'image" revue RIA-ECA vol 17, Ed.Hermès, ISBN 2-7462-0631-5, Janvier 2003, pp.219-231.
5. M. Hammami, Y. Chahir, L. Chen, "Combining Text and Image Analysis in The Web Filtering System: WebGuard", *IADIS International Conference: WWW/Internet 2003*, ISBN 972-98947-1-X, Algarve, Portugal, November 5-8, pp.611-618
6. M.J. Jones, J.M. Regh, “Statistical Color Models with application to Skin Detection”, Cambridge Research Laboratory, CRL 98/11, 1998.
7. J. R. Quinlan. Induction of decision trees. *Machine Learning*, 1:81–106, 1986.
8. J. R. Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufmann, 1993.
9. D. Tock and I. Craw. Tracking and measuring drivers’ eyes. *Image and Vision Computing*, 14:541–548, 1996.
10. H. Wang and S-F. Chang, “A highly efficient system for automatic face region detection in mpeg video,” *IEEE Transactions on circuits and system for video technology*, vol. 7, no. 4, pp. 615–628, August 1997.
11. J. G. Wang and E. Sung, “Frontal-view face detection and facial feature extraction using color and morphological operators,” *Pattern recognition letters*, vol. 20, no. 10, pp. 1053–1068, October 1999.
12. Ming-Hsuan Yang and Narendra Ahuja, “Detecting human faces in color images,” in *Proceedings of the International Conference on Image Processing*, Chicago, IL, October 4-7 1998, pp. 127–130.
13. D.A. Zighed and R. Rakotomala. A method for non arborescent induction graphs. Technical report, Laboratory ERIC, University of Lyon 2, 1996.