Face Detection Using a Dual Cross-Validation of Chrominance/Luminance Channel Decisions and Decorrelation of the Color Space

VICTOR-EMIL NEAGOE, MIHAI NEGHINĂ
Depart. Electronics, Telecommunications & Information Technology
Polytechnic University of Bucharest
Splaiul Independentei No. 313, Sector 6, Bucharest
ROMANIA
E-mail:victoremil@gmail.com

Abstract: - We propose a new face detection model based on the competition between the chrominance and luminance channel decisions. Each of the two detection branches has its own techniques of finding face candidates and the model implies a dual cross-validation of the above channels. One investigates the decision improvement of skin detection over the color channel by applying the conversion from the conventional RGB space into the 3D uncorrelated color space (UCS), using the Karhunen-Loève transform (KLT) in the color space. One evaluates the performances of the proposed model using an UCS by comparison to other two well known color representation (YCbCr) and (HSV). For experimental evaluations, we have chosen 120 images from "Labeled Faces in the Wild" database. The proposed algorithm leads to a correct detection score with about 7% better than the classical Viola-Jones method. The detection rates obtained using UCS representation are better by comparison to YCbCr and HSV color spaces.

Key-Words: - Detection, Cross-Validation, Chrominance/Luminance Channels, Uncorrelated Color Space

1 Introduction

Face detection from images is a key problem in human computer interaction studies and in pattern recognition researches. Detecting faces is the first crucial step for face recognition and as a consequence it has significant applications such video surveillance, human computer interface, and face image database management. Many different approaches for face localization are published in the literature.

All the face detection attempts rely on only one type of algorithm for detection, sometimes followed by other types of techniques for further validation. Most methods use only the luminance component, extracting features as texture, depth, shape, and eigenfaces. There are applied various techniques as learning algorithms, bootstraps, SVM, neural networks and fuzzy methods.

A second group of methods added the chrominance information as a validation of the luminance channel technique. The general idea behind them was to confirm through color analysis that the candidate object has face-like color.

A third type of algorithms starts from the chrominance information to locate candidate faces, which are then validated by searching other facial features. The necessity of finding suitable facial features for validation is due to the fact that color analysis yields information related to the presence of skin rather than the presence of face.

In this paper, we propose a model that belongs to a new algorithm category consisting of the two competing luminance/ chrominance channels. Our model uses two corresponding techniques of finding face candidates and proposes a dual cross-validation of the two branch decisions in an attempt to cover different situations. This cross-validation is different from the simple logical "AND" of the two channel decisions. The color channel uses skin detection based on color histogram, followed by analysis of shape information and ellipse fitting. When the luminance channel has the function of main face detector, it applies the Viola-Jones algorithm [11]. For validation of the chrominance detection, the luminance channel applies an improved SVM technique for fast detection [4], [8].

When building a system that uses skin color as a feature for face detection, first main problem is to choose the color space. The main purpose of this paper is dedicated to the improvement of face detection rate in color images by applying the color space conversion model previously proposed by Neagoe in [6], [7] for color pattern recognition; this implies the color space conversion from the conventional RGB space into the 3D *uncorrelated color space* (UCS), using the Karhunen-Loève transform (KLT), equivalent to Principal Component Analysis (PCA) in the color space.

The paper is structured as follows. Second section presents the flowchart of the proposed algorithm. Third section is dedicated to the experimental results, while the fourth section contain concluding remarks.

2 Model Description

The present model has three main sources. Firstly, for chrominance channel, this model is inspired by the approach of Sobottka and Pitas [10] for face detection, based on the observation that human faces are characterized by their oval shape and skin-color, also in the case of varying light conditions. It applies color segmentation in HSV color space and it is followed by analysis of the shape information by ellipse fitting. Secondly, for luminance channels, this paper applies either Viola-Jones face detection algorithm [11] or the fast SVM of Kienzle *et al* [4]. Thirdly, to the aim of improving detection performance, we have applied the color space model proposed by Neagoe [6], [7] consisting in conversion of the conventional RGB space into the 3D *uncorrelated color space* (UCS), using the Karhunen-Loève transform (KLT).

On the other side, face detection algorithms presented in literature belong to one of the following three categories: (a) methods using luminance information; (b) algorithms adding the chrominance decision as a validation of the luminance decision; (c) methods based only on the chrominance information for face candidate detection.

We further propose a face detection model (see Fig. 1), thus initiating a new category. It has two competing chrominance/luminance branches and two specific face candidate finding techniques that cross-validate their decisions each other; the final decision is a result different of a simple logical intersection of branch decisions. The main branch of the flow-chart corresponds to the locating of the candidates by chrominance analysis and to the cross-validation by a technique relying on the luminance channel. We have covered several color spaces, to comparatively evaluate the algorithm detection performance.

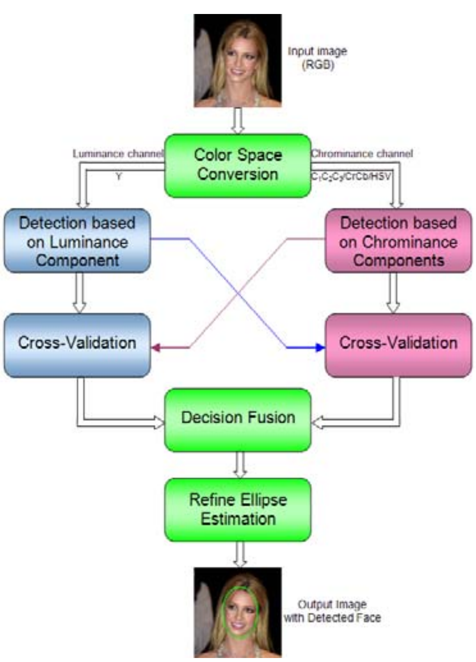

Fig. 1. Flow-chart of the proposed face detection algorithm.

2.1 Color Space Conversion

We further consider the Karhunen-Loève transformation (KLT) for conversion of the RGB space into a 3D *Uncorrelated Color Space (UCS)*. For comparison, one evaluates the application of the YCbCr and HSV color spaces.

Decorrelation of the Color Space. In order to improve the color pattern recognition performances, we shall further apply the color space conversion model proposed by Neagoe in [6], [7]. Consider the color pixels in a given image as 3D vectors

$$P(x, y) = [R(x, y)\ G(x, y)\ B(x, y)]^t \qquad (1)$$

where R(x, y), G(x, y) and B(x, y) are the red, green and blue components of the pixel of coordinates (x, y).

We assume that color images exhibit features that can be useful in the conversion from a 3D full color space representation to the 3D uncorrelated color space (UCS); as we further prove, this transformation improves correct pattern recognition score. For color conversion, we have chosen the Karhunen-Loève transformation (KLT), also known as Principal Component Analysis (PCA), to eliminate the correlation of the R, G, and B color channels. The interesting fact in our case is that one preserves the space dimensionality (from 3D RGB to 3D UCS). The reason of applying PCA here is not to reduce the space dimension but only to eliminate the component correlation.

To deduce the KLT matrix, one firstly computes the covariance matrix of the color pixels (represented as 3D vectors). Then, one computes the eigenvalues of the covariance matrix. Finally, we deduce the three eigenvectors. Thus, one obtains the KLT matrix K

$$K = \begin{bmatrix} A^t \\ B^t \\ C^t \end{bmatrix}, \text{ with } A = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}, B = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}, C = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}, \qquad (2)$$

where A, B, and C are the eigenvectors of the covariance matrix, and t denotes trans-position.

Then, the transformation of the 3D color vector P(x, y) into the 3D UCS is the vector C(x,y)

$$C(x, y) = \begin{bmatrix} C_1(x,y) \\ C_2(x,y) \\ C_3(x,y) \end{bmatrix} \qquad (3)$$

given by the equation

$$C(x, y) = K \bullet P(x, y) \qquad (4)$$

Example 3.1. By selecting a set of pictures containing skin regions, we have computed the co-variance matrix of 3D color space RGB, and then the corresponding eigenvalues and eigenvectors. Thus, one deduces the matrix defining the Karhunen-Loève Transform (KLT), so that for this particular case, the color space transformation is given by the following equations

$$\begin{aligned} C1 &= \ \ 0.6369 \bullet R + 0.5993 \bullet G + 0.4848 \bullet B \\ C2 &= -0.6852 \bullet R + 0.1520 \bullet G + 0.7122 \bullet B \\ C3 &= -0.3531 \bullet R + 0.7859 \bullet G - 0.5074 \bullet B \end{aligned} \qquad (5)$$

As a result of space conversion using the matrix K, the covariance matrix in the UCS is a diagonal matrix whose elements approximate the eigenvalues of the covariance matrix of the RGB space matrix. One can remark that all the off-diagonal elements of the transformed covariance matrix became zero, corresponding to the definition of the uncorrelated color space.

2.2 Detection based on Chrominance Components

Skin Segmentation. The aim of skin color pixel classification is to determine if a color pixel is a skin color or non skin color. Good skin color pixel classification should provide coverage of all different skin types (blackish, yellowish, brownish, whitish, etc.). We have evaluated the performances of skin color pixel classification for the above mentioned three color spaces: UCS (C1C2C3), YCbCr, and HSV.

UCS (C1C2C3). The skin detection rule is given by the following set of inequations

$\alpha_{UCS} < C1 < \beta_{UCS}; \gamma_{UCS} < C2 < \delta_{UCS}; \lambda_{UCS} < C3 < \theta_{UCS}$ (6)

where the thresholds α_{UCS}, β_{UCS}, γ_{UCS}, δ_{UCS}, λ_{UCS} and θ_{UCS} are estimated from the histogram of the UCS components. For a set of pictures including skin regions we have obtained the histograms of C1, C2, C3, given in Fig. 2. One can remark that we can neglect the first inequation regarding C1.

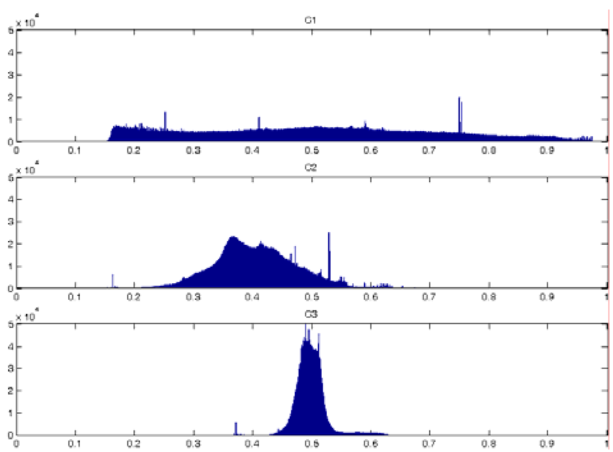

Fig. 2. Histogram of the UCS components (C1, C2, C3).

YCbCr. The skin detection rule is expressed by the following two inequations

$\alpha_{YCbCr} < Cb < \beta_{YCbCr}; \gamma_{YCbCr} < Cr < \delta_{YCbCr}$ (7)

From the CbCr component histogram, one can deduce the corresponding threshold parameters α_{YCbCr}, β_{YCbCr}, γ_{YCbCr}, δ_{YCbCr}.

HSV. The inequations corresponding to skin detection are

$\alpha_{HSV} < H < \beta_{HSV}; \gamma_{HSV} < S < \delta_{HSV}; \lambda_{HSV} < V < \theta_{HSV}$(8)

where the threshold parameters α_{HSV}, β_{HSV}, γ_{HSV}, δ_{HSV}, λ_{HSV} and θ_{HSV} are estimated from the HSV component histogram. We can remark that we can neglect the inequation regarding V.

Region Growing. The reason of considering region growing is to increase robustness against noise and changes in illumination. The connected components are determined by applying a region growing algorithm at a coarse resolution of the segmented image. We have considered a structured element of 5x5 pixels in order to fill the holes in the extracted skin regions.

Ellipse Estimation. The oval shape of a face can be approximated by an ellipse. This is why looking for faces in an image means to detect objects with nearly elliptical shape. After region growing for each connected component with a given minimum size, the best-fit ellipse is computed according to the paper of Sobottka and Pitas, where the ellipse estimation is given for the particular case of the HSV space [10]. We have estimated the following region parameters: area, centroid, bounding box, and ellipse parameters (orientation and lengths of major and minor axes).

2.3 Luminance Cross-Validation of the Chrominance Detection

For the chrominance branch, one applies a cross-validation of the detected candidates using the luminance as an independent channel, based on the Support Vector Machine (SVM) classifier, recognized as a performing technique. However, SVMs are considered slower at runtime than other methods with similar generalization performance. We have applied the technique developed by Kienzle *et al* [4] for reducing SVM runtime complexity; the set of support vectors is replaced by a smaller, so-called reduced set of synthesized input space points. Before validation, one selects and extracts the most fitted bounding box (with most skin pixels and ellipse approximated shape). This bounding box is rotated in four positions, each position having a vertical major or minor ellipse axis. One checks SVM luminance detection of at least one of these positions; one sets the SVM detection level for validation at a low value.

2.4 Detection based on Luminance Component

For face candidate detection over the luminance channel, we have applied the Viola-Jones algorithm [10], that is capable of processing images extremely rapidly while achieving high detection rates. It uses a simple and efficient classifier which is built using the AdaBoost learning algorithm to select a small number of critical visual features from a very large set of potential features. The Viola-Jones method combines classifiers in a "cascade" which allows background regions of the image to be quickly discarded while spending more computation on promising face-like regions.

2.5 Chrominance Cross-Validation of the Luminance Detection

The face candidates detected by Viola-Jones luminance channel are submitted to a cross-validation by the color skin detection corresponding to the chrominance branch. The chrominance validation ensures that the detected face candidate contains a significant number of pixels of skin-like color.

2.6 Decision Fusion

The final decision has as aim to avoid detection overlapping as well as avoiding multiple detections for the same face. The decision fusion is controlled by the following rules:

– *Detection branch*. Chrominance channel detection validated by luminance has the priority, followed by the luminance branch detection validated by chrominance.
– *Candidate size*. The bigger face candidate has priority.

For example, by combining the above two decision rules, the first position is assigned to the bigger face candidate detected by chrominance channel and validated by luminance branch, the second position is given to the second size candidate detected by chrominance and validated by luminance, then the third position is allocated to the candidate detected by luminance and validated by chrominance.

2.7 Refining of the Ellipse Estimation

Refining the parameters of the ellipse means a more accurate face approximation, starting from the rectangle provided at the output of the luminance SVM. For example, we have approximated the two ellipse axes, started from the sizes of the above mentioned rectangle.

3 Experimental Results

3.1 "Labeled Faces in the Wild" Database

To evaluate the performances of the proposed model, we have selected pictures from the database "Labeled Faces in the Wild" [3].

This database of face photographs has been designed by the UC Berkeley for studying the problem of unconstrained face recognition [3]. The dataset contains more than 13,000 images of faces (250 x 250 pixels) collected from the web. A number of 1680 subjects pictured have two or more distinct photos in the data set. For our experimental evaluation, we have selected 120 images belonging to 30 subjects (four pictures for each subject) as shown in Fig. 3.

Fig. 3. Examples of selected pictures from the "Labeled Faces in the Wild" Database.

3.2 Examples of images processed by the proposed algorithm

In Fig. 4 one can see an example of images provided by the proposed processing cascade for the case of face detection by chrominance branch validated by luminance channel. However, this face is also detected by the luminance channel, but the system applies the priority rule of decision fusion mentioned in 2.6.

Fig. 4. Example of face detected by the chrominance branch and validated by the luminance channel (it is also detected by the luminance channel): (a) original image; (b) skin color segmented image; (c) connected components; (d) luminance detection result; (e) best-fit ellipse.

In Fig. 5 we show output images of the processing cascade for the case of face detection by the luminance branch and validated by the chrominance channel. It is a special case where the first branch (chrominance channel) fails to detect faces since the background has a skin-like color. This can also happen when the skin region of the face is connected to other skin regions (e.g. dresses that show the neck, shoulders, hands connected, hands or other faces touching faces, thus completely confusing the ellipse-fitting step).

Fig. 5. Example of face detected by the luminance branch and validated by the chrominance channel (but it is not detected by the chrominance channel!): (a) original image; (b) skin color segmented image; (c) connected components; (d) luminance detection result; (e) best-fit ellipse.

Fig. 6 shows a special case where the first branch (chrominance) correctly detects faces but the second branch (luminance) fails to detect faces; however, the luminance channel validates the chrominance detection. Such cases correspond to images where the face is rotated.

Fig. 6. Example of face detected by the chrominance branch and validated by the luminance channel (but it is not detected by the luminance channel!): (a) original image; (b) skin color segmented image; (c) connected components; (d) luminance detection result (failure of detection); (e) best-fit ellipse.

3.3 Performance Evaluation

In order to evaluate the performances of the proposed face detection algorithm, we have chosen the following parameters:

– *TP* = true positives = number of correctly detected faces

– *FN* = false negatives = number of lost faces

– *FP* = false positives = number of incorrectly detected items as faces

– *T* = total of existing faces = *TP* + *FN* (we have one face per image)

– *CTP* = chrominance based true positives = number of correctly detected faces by the chrominance channel and validated by luminance channel

– *LTP* = luminance based true positives = number of correctly detected faces by the luminance channel and validated by chrominance channel

– $TP = CTP + LTP$

– CDR = correct detection rate = $\frac{TP}{T} \cdot 100 \;\; [\%]$

– $CCDR$ = chrominance correct detection rate = $\frac{CTP}{TP} \cdot 100 \;\; [\%]$

– $LCDR$ = luminance correct detection rate = $\frac{LTP}{TP} \cdot 100 \;\; [\%]$

– FPR = false positive rate = $\frac{FP}{T} \cdot 100 \;\; [\%]$

– MR = missing rate = $\frac{FN}{T} \cdot 100 \;\; [\%]$

To evaluate the performances of our model, we have selected 120 pictures from "Labeled Faces in the Wild" database. The experimental results are given in Tables 1 and 2 proving both the advantage of the new algorithm over Viola-Jones method, as well as the advantage of the proposed uncorrelated color space (UCS) representation over {Cb,Cr} and {H,S,V} color spaces.

Table 1. Performances of the proposed face detection algorithm (the test lot has T=120 images selected from „Labeled Faces in the Wild" database).

Color space variant	TP	CTP	LTP	FP	FN	CDR [%]		FPR [%]
						CCDR [%]	LCDR [%]	
UCS={(C1C2C3)}	106	59	47	22	21	**88.33**		18.33
						55.67	44.33	
{Cb,Cr}	105	19	86	21	15	87.50		17.50
						20	80	
{H,S,V}	104	42	62	23	16	86.67		19.17
						40.38	59.62	

Table 2. Performances of the Viola-Jones face detection algorithm (the test lot has T=120 images selected from „Labeled Faces in the Wild" database).

TP	FP	FN	CDR [%]	FPR [%]
98	22	22	**81.67**	18.33

4 Concluding Remarks

1. The paper proposes a new face detection model. The algorithm has two competing branches: chrominance and luminance channels. Each channel has its specific detection technique. The model implies a dual cross-validation. Each channel detection is cross validated by the other one. This has the benefit of covering more cases for face detection in complex background color images, but also has the disadvantage of requiring a greater amount of computational power compared to more straightforward algorithms.

2. When the luminance channel is used for primary detection of face candidates, we have applied the fast and accurate Viola-Jones algorithm; when the luminance branch is applied for cross-validation of the chrominance candidates, it uses the Support Vector Machine (SVM) classifier.

3. In order to improve the chrominance face detection performances, one proposes the RGB space conversion into the 3D uncorrelated color space (UCS) for skin detection. This method introduced in [6], [7] for color face recognition is applied now for face detection. It is based on the Karhunen-Loève transform (KLT) equivalent to PCA in the color space. The reason of applying KLT here is not to reduce the space dimension, but to increase the skin detection score by removing the color component correlation.

4. For the considered database, the proposed algorithm leads to a correct detection score (CDR) of 88.33 %, with about 7% better than the classical Viola-Jones method (see Tables 1 and 2).

5. We have evaluated the face detection performances when the algorithm uses the skin detection in the UCS, by comparison to face detection when skin detection uses the YCbCr and HSV color spaces. We have pointed out the advantage of face detection using UCS. One remarks that the advantage of decorrelating the color space (applying UCS) can not be pointed out completely here, since skin segmentation involving the color space is only a part of the complex face detection algorithm.

References:

[1] D. Chai, K.N. Ngan, Face Segmentation Using Skin Color Map in Videophone Applications, *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 9. No. 4, 1999, pp. 551-564.

[2] R.L. Hsu, M.A. Mottaleb, A.K. Jain, Face Detection in Color Images, *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 24. No. 5, 2002, pp. 696-705.

[3] G.B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*, University of Massachusetts, Amherst, Technical Report No. 07-49, 2007.

[4] W. Kienzle, G. Bakır, M. Franz, B. Schoelkopf, Face Detection - Efficient and Rank Deficient, *Advances in Neural Information Processing Systems*, Vol. 17, 2005, MIT Press, Cambridge, pp. 673-680.

[5] B. Menser, F. Mueller, Face Detection in Color Images Using Principal Components Analysis, *Proc. Seventh International Conference on Image Processing and Its Applications,* Vol. 2., 1999, pp. 620-624.

[6] V.E. Neagoe, An Optimum 2D Color Space for Pattern Recognition, *Proc. of the 2006 International Conference on Image Processing, Computer Vision&Pattern Recognition (ICPV'08)*, *WORLDCOMP'06,* Las Vegas, Vol. 2, 2006, pp. 526-532.

[7] V.E. Neagoe, Decorrelation of the Color Space, Feature/Decision Fusion, and Concurrent Neural Classifiers for Color Pattern Recognition, *Proc. 2008 International Conference on Image Processing, Computer Vision & Pattern Recognition (ICPV'08)*, *WORLDCOMP'08*, Las Vegas, pp. 28-34.

[8] E. Osuna, F. Girosi, Reducing the Run-Time Complexity in Support Vector Machines. *Advances In Kernel Methods–Support Vector Learning*, 1999, MIT Press, Cambridge, pp. 271–284

[9] S.L. Phung, A. Bouzerdoum, D. Chai, Skin segmentation using color pixel classification: analysis and comparison, *IEEE Trans. Patt. Anal. Mach. Intell.,* Vol. 27. No. 1, 2005, pp. 148-154

[10] K. Sobottka, I. Pitas, Face Localization and Facial Feature Extraction based on Shape and Color Information, *Proc. International Conference on Image Processing,* Vol. 3, 1996, pp. 483-486.

[11] P. Viola, M. Jones, Robust Real-Time Face Detection, *International Journal of Computer Vision,* Vol. 57, 2004, nr. 2, pp. 137-154.

[12] M.H. Yang, D. Kriegman, N. Ahuja, Detecting Faces in Images: A Survey, *IEEE Trans. Patt. Anal. Mach. Intell.,* Vol. 24. No. 1, 2002, pp. 34-55.