HUMAN FACE DETECTION IN VIDEO USING EDGE PROJECTIONS

Mehmet Türkan, Ibrahim Onaran, and A. Enis Çetin

Department of Electrical and Electronics Engineering, Bilkent University
Bilkent, TR-06800, Ankara, Turkey
phone: + (90) 312 290-1477, fax: + (90) 312 266-4192, email: {turkan, onaran, cetin}@ee.bilkent.edu.tr
web: www.ee.bilkent.edu.tr

ABSTRACT

In this paper, a human face detection algorithm in images and video is presented. After determining possible face candidate regions using colour information, each region is filtered by a high-pass filter of a wavelet transform. In this way, edges of the region are highlighted, and a caricature-like representation of candidate regions is obtained. Horizontal, vertical, filter-like and circular projections of the region are used as feature signals in support vector machine (SVM) based classifiers. It turns out that our feature extraction method provides good detection rates with SVM based classifiers.

1. INTRODUCTION

Human face detection problem has received significant attention during the past several years because of wide range of commercial and law enforcement applications. In recent years, many heuristic and pattern recognition based methods have been proposed to detect human faces in still images and video on gray-scale or colour. Human face detection techniques based on neural networks, support vector machines, hidden Markov models, Fisherspace/subspace linear discriminant analysis, principle component analysis, and Bayesian or maximum-likelihood classification methods have been described in the literature ranging from very simple algorithms to composite high-level approaches. Rowley et al. [1] presented a neural-network based upright frontal face detection system using a bootstrap algorithm. The performance over a single network is improved by arbitrating the system among multiple networks. Sung and Poggio [2] described a distribution-based modelling of face and non-face patterns using a multilayer perceptron (MLP) classifier. They developed a successful example-based learning system for detecting vertical frontal views of human faces in complex scenes. Osuna et al. [3] demonstrated a decomposition algorithm to train support vector machines for frontal human face detection in images over large data sets. Nefian and Hayes [4] described a hidden Markov model (HMM)-based framework using the projection coefficients of the Karhunen-Loeve Transform (KLT) for detection and recognition of human faces. Turk and Pentland [5] presented an eigenface based human face recognition technique using principal component analysis. They developed a system which tracks the head of a subject and then the person is recognized by comparing characteristics of the face to those of known individuals with a nearest-neighbour classifier. Conceptually detailed literature survey on human face detection is conducted by Hjelmas and Low [6].

Recently, wavelet based face detection methods have been developed and become very popular. The main reason is that a complete framework has been recently built in particular for what concerns the construction of wavelet bases and efficient algorithms for the wavelet transform computation [7]. Wavelet packets allow more flexibility in signal decomposition and dimensionality reduction as the computational complexity is an important subject for face detection systems. Garcia and Tziritas [7] proposed a wavelet packet decomposition method on the intensity plane of the candidate face regions. After obtaining the skin colour filtered (SCF) image using the colour information of the original image, they extracted feature vectors from a set of wavelet packet coefficients in each region. Then, the face candidate region is classified into either face or non-face class by evaluating and thresholding the Bhattacharyya distance between the candidate region feature vector and a prototype feature vector. Zhu et al. [8] described a subspace approach to capture local discriminative features in the space-frequency domain for fast face detection based on orthonormal wavelet packet analysis. They demonstrated the detail (high frequency) information within local facial areas. For example; eyes, nose, and mouth show noticeable discrimination ability for face detection problem of frontal view faces in a complex background. The algorithm leads to a set of wavelet features with maximum class discrimination and dimensionality reduction. Then the classification is evaluated by a likelihood test. Uzunov et al. [9] described an adequate feature extraction method in a face detection system. The optimal atomic decompositions are selected from various dictionaries of anisotropic wavelet packets using the adaptive boosting algorithm (AdaBoost) [10]. Their method demonstrates a fast learning process with high detection accuracy.

In this study, a human face detection method in images and video is proposed on both gray-scale and colour. Our method is based on the idea that a typical human face can be recognized from its edges. In fact, a caricaturist draws a face image in a few strokes by drawing the major edges of the face. Most wavelet domain image classification methods are also based on this fact because wavelet coefficients are closely related with edges [7].

After determining all possible face candidate regions in a given video frame or still image, each region is single stage 2-D rectangular wavelet transformed. In this way, wavelet domain sub-images are obtained. The low-high and high-low sub-images contain horizontal and vertical edges of the region, respectively. The high-high sub-image may contain almost all the edges, if the face candidate region is sharp enough. It is clear that, the detail information within local facial areas, e.g., eyes, nose, and mouth, show noticeable discrimination ability for face detection problem of frontal view faces [8]. We take the advantage of this fact by summarizing these sub-images using their projections, and obtain 1-D projection feature vectors corresponding to edge images of face or face-like regions. The advantage of the projections is that they can be easily normalized to a fixed size and this provides robustness against scale changes. The horizontal and vertical projections are simply computed by summing the absolute pixel values in a row and column in a given edge image, respectively. Furthermore, filter-like projections are computed as in Viola and Jones [10] approach as additional feature vectors. The final feature vector for a face candidate region is obtained by combining all the horizontal, vertical, and filter-like projections. These feature vectors are then classified using support vector machine (SVM) based classifiers into face or non-face classes.

This paper is organized as follows. Section 2 specifies a general block diagram of our face detection system where each block is briefly described for the techniques used in the implementation. In Section 3, a short introduction to face detection problem for omnidirectional camera view is described. In Section 4, the detection performance of support vector machines is compared with currently available face detection methods. Conclusions are also presented in Section 4.

2. FACE DETECTION SYSTEM

In this paper, a human face detection scheme for frontal pose and upright orientation is developed (Fig. 1). After determining all possible face candidate regions in a given video frame or still image, each region is decomposed into its wavelet domain sub-images as shown in Fig. 2. Face candidate regions can be estimated based on color information in video as described in Section 2.1. The detail information within local facial areas, e.g., eyes, nose, and mouth, is obtained in high-high sub-image of the face pattern. This sub-image is similar to a hand-drawn face image, and in a given region, face patterns can be discriminated using this high-pass filtered sub-image. Other high-band sub-images can be also used to enhance the high-high sub-image. The wavelet domain processing is presented in Section 2.2. For a face candidate region, a feature vector is generated from wavelet domain sub-images using projections as explained in Section 2.3. The generated feature vectors are then classified using support vector machine (SVM) based classifiers into face or non-face classes. The detailed description of these classifiers is given in Section 2.4.

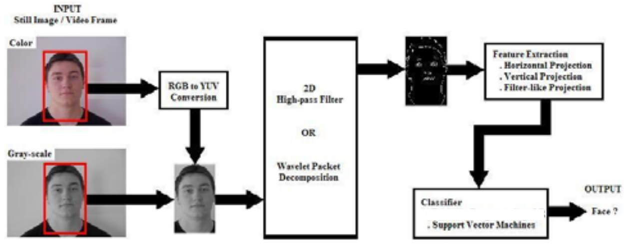

Figure 1 – Block diagram of the face detection system.

2.1 Detection of face candidate regions

In this study, especially for real-time implementation, the RGB colour space with pixel-based skin detection method is chosen for fast processing. Given a colour video frame or still image, each pixel is labelled as skin or non-skin using a number of predefined rules [12]. Then morphological operations are performed on skin labelled pixels in order to have connected face candidate regions. The candidate regions' intensity images are then fed into a 2-D high-pass filter or a single stage 2-D rectangular wavelet decomposition block.

2.2 Wavelet decomposition of face patterns

Possible face candidate regions are processed using a two-dimensional filterbank. The regions are first processed row-wise using a 1-D filterbank with a low-pass and high-pass filter pair, h and g, respectively. Resulting two image signals are processed column-wise once again using the same filterbank. The high-band sub-images that are obtained using a high-pass filter contain edge information, e.g., the low-high and high-low sub-images contain horizontal and vertical edges of the input image, respectively (Fig. 2). Therefore, absolute values of low-high, high-low and high-high sub-images can be added to have an image having significant edges of the candidate region. We call this image the *detail* image. Lagrange filterbank [13] consisting of the low-pass filter, $h[n]=\{0.25, 0.5, 0.25\}$, and the high-pass filter $g[n]=\{-0.25, 0.5, -0.25\}$ is used in this paper.

A second approach is to use a 2-D low-pass filter and subtract the low-pass filtered image from the original image. The resulting image also contains the edge information of the original image and it is equivalent to the sum of undecimated low-high, high-low, and high-high sub-images.

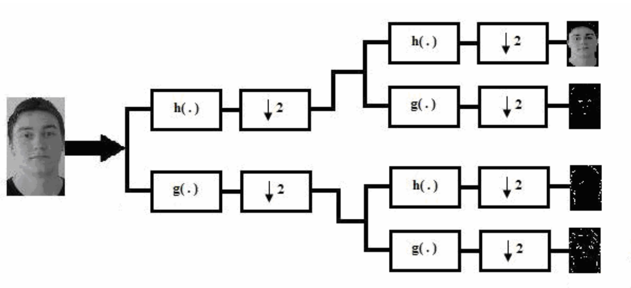

Figure 2 – Two-dimensional rectangular wavelet decomposition of a face pattern; low-low, low-high, high-low, high-high sub-images. h and g represent 1-D low-pass and high-pass filters, respectively.

2.3 Feature extraction

In this paper, the edge projections of the candidate image regions are used as features. Edge information of the original image is available obtained using the wavelet analysis. The components of the feature vector are the horizontal, vertical, and filter-like projections of the wavelet subimages. The advantage of the 1-D projection signals is that they can be easily normalized to a fixed size and this provides robustness against scale changes.

Horizontal projection $H(.)$, and vertical projection $V(.)$ are simply computed by summing the absolute pixel values, $d(.,.)$, of the *detail* image in a row and column, respectively as follows,

$$H(y)=\sum_{x}\left|d(x,y)\right|$$

$$V(x)=\sum_{y}\left|d(x,y)\right|$$

In this way, we take the advantage of the detail information within local facial areas, e.g., eyes, nose, and mouth, in both horizontal and vertical directions. These two projections actually provide us significant discrimination ability for classification. Typical projection vectors after low-pass smoothing and normalization are shown in Fig. 4.

Furthermore, filter-like projections, $F_i(.)$, are computed similar to Viola and Jones [10] approach as additional feature vectors. We divide the *detail* image into two regions, R_1 and R_2, as shown in Fig. 3, and compute projections in these regions. We subtract the horizontal projections in R_1 and R_2, and obtain a new horizontal projection vector $F_1(y)$. In this way, the symmetry property of a typical human face is considered as a feature also.

$$F_1(y)=\left|\sum_{x\in R_1}\left|d(x,y)\right|-\sum_{x\in R_2}\left|d(x,y)\right|\right|$$

Because of the symmetry property of a face pattern, especially vertical-cut filter-like projections are very close to zero. Similarly, a new vertical projection vector $F_2(x)$ is computed as follows,

$$F_2(x)=\left|\sum_{y\in R_1}\left|d(x,y)\right|-\sum_{y\in R_2}\left|d(x,y)\right|\right|$$

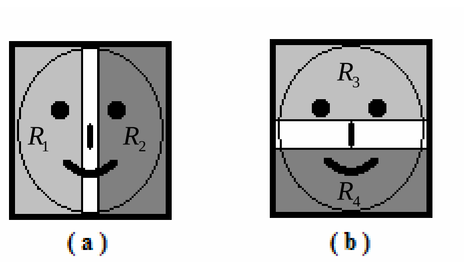

Figure 3 – Two-rectangle feature regions. White area between the feature regions is a “dead-zone” in which no summation is carried out; (a) vertical-cut, (b) horizontal-cut of face candidate regions.

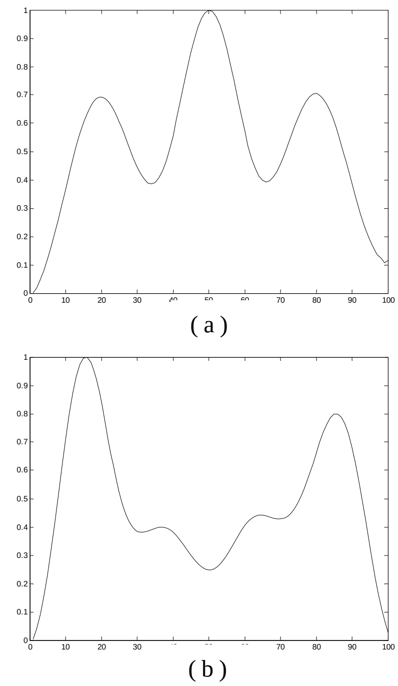

Figure 4 –A typical normalized (a) horizontal, (b) vertical projection. In horizontal projection, the first peak corresponds to the left eye, second peak corresponds to nose and mouth, and the third peak corresponds to the right eye, respectively.

Projection vectors are concatenated to obtain a composite feature vector. A composite feature vector consisting of the projections $H(.)$, $V(.)$, $F_1(.)$ and $F_2(.)$ are used to represent a given image region.

We also repeat this process for regions R_3 and R_4, shown in Fig. 3, and obtain additional feature vectors.

2.4 Support vector machines (SVMs)

After generating feature vectors, the face detection problem is reduced to a classification problem. In this paper, the approach that we studied is support vector machines (SVMs), which are a brand new and powerful machine learning technique, based on structural risk minimization for both regression and classification problems. An important benefit of the support vector machine approach is that the complexity of the resulting classifier is characterized by the number of support vectors rather than the dimensionality of the transformed space [14]. Thus, SVMs compensate the problems of overfitting unlike some other classification methods.

Support vector machines seek to define a linear boundary between classes such that the margin of separation between samples from different classes that lie next to each other is maximized. Classification by SVMs is concerned only with data from each class near the decision boundary, called support vectors. Support vectors lie on the margin and carry all the relevant information about the classification problem. They are informally the hardest patterns to classify, and the most informative ones for designing the classifier [14].

This approach is generalized to non-linear case by mapping the original feature space into some other space using a mapping function and performing optimal hyperplane algorithm in this dimensionally increased space. In the original feature space, the hyperplane corresponds to a non-linear decision function whose form is determined by the mapping kernel. Mapping kernels have been developed to compute the boundary as a polynomial, sigmoid, and radial basis function (RBF).

In this paper, we used a library for support vector machines called LIBSVM [15] which is available online for free of charge in either C++ or Java, with interfaces for Matlab, Perl and Python. Our simulations are carried out in C++ with interface for Python using radial basis function (RBF) as kernel with default parameter selection. RBF kernels are computed as follows,

$$k(x,y)=e^{\frac{-\|x-y\|^2}{2\sigma^2}}$$

LIBSVM package provides the necessary quadratic programming routines to carry out classification. It also normalizes each feature by linearly scaling it to the range $[-1,+1]$, and performs cross validation on the data set.

Support vector machines are used for isolated handwritten digit detection, object recognition, and also face detection by several researchers, i.e., Osuna et al. [3].

3. OMNIDIRECTIONAL FACE DETECTION

In this paper, we also studied human face detection problem in omnidirectional camera view. Unlike previous sections, this section is related with distorted and oriented faces in both vertically and horizontally in real-time. A sample of omnidirectional camera view, and samples of human faces captured from an omnidirectional camera are presented in Fig. 5.

After detecting possible face candidate regions using colour information, each region is decomposed into its sub-images in the wavelet domain. Absolute values of wavelet images are added, and the resulting image is thresholded to obtain binary images as shown in Fig. 5. Then the feature vectors are extracted using circular feature regions as shown in Fig. 6. Circular projections are simply computed by summing the absolute pixel values of the *detail* image in a given circular region. Furthermore, as a Haar-like feature, we multiply bold shaded circular regions with -1, and subtract the result from its adjacent light shaded circular region and obtain an additional feature vector. Typical circular projection vectors after 1-D low-pass smoothing and normalization are shown in Fig. 7.

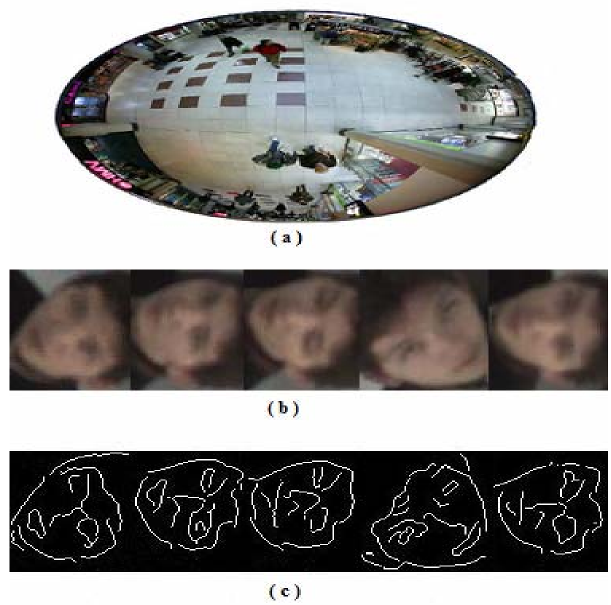

Figure 5 – (a) A sample view of omnidirectional camera, which is taken from Grandeye Company in UK, (b) human face samples captured from an omnidirectional camera, (c) binary images of face samples shown in (b).

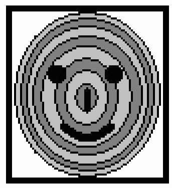

Figure 6 – Circular feature regions for omnidirectional cameras.

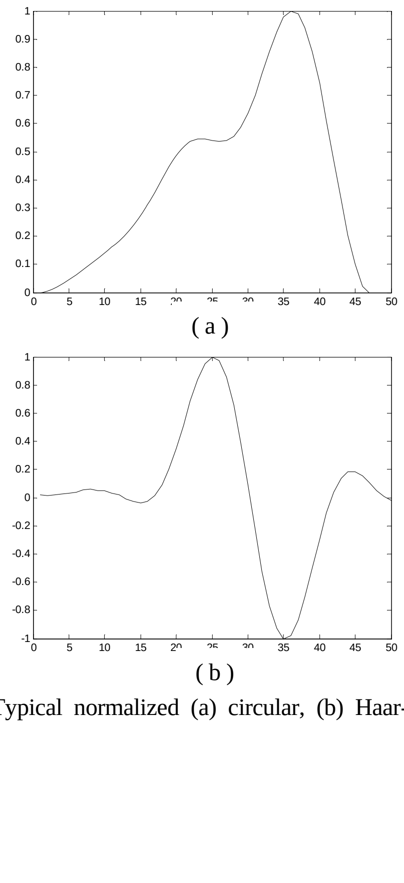

Figure 7 –Typical normalized (a) circular, (b) Haar-like circular projections.

4. RESULTS AND CONCLUSIONS

The proposed face detection algorithm was evaluated on several face image databases including the Computer Vision Laboratory (CVL) Face Database [16]. The database contains 797 color images of 114 persons. Each person has 7 different images of size 640x480 pixels; far left side view, 45^o angle side view, serious expression frontal view, 135^o angle side view, far right side view, smile -showing no teeth- frontal view, and smile -showing teeth- frontal view. We extracted 335 frontal view face dataset from this database by cropping faces of variable size. Furthermore, 100 non-face samples are extracted from color images downloaded randomly from the World Wide Web. These non-face samples include skin colored objects and typical human skin regions such as hands, arms, and legs. The best success rate achieved using SVMs is 99.6% over whole face and non-face datasets with concatenation of horizontal, vertical, and vertical-cut filter-like projections as feature vectors. While training SVMs, 100 face samples and 50 non-face samples used, and then these are also included in test set.

The second experimental setup consists of a real-time human face dataset for real-time implementation. This dataset is collected in our laboratory. The dataset currently contains the video of 12 different people with 30 frames each. A person's face is recorded from 45^o side view to 135^o side view from different distances to camera with a neutral facial expression under the day-light illumination. Then, SVMs are trained with these data and the resulting modal file of LIBSVM is used for classifying the test features in real-time. The proposed human face detection system is implemented in .NET C++ environment, and it works in real-time with 15 fps on a standard personal computer. We used a Philips web camera with output resolution of 320x240 pixels throughout all our real-time experiments.

The third database that we used in our experiments is the Facial Recognition Technology (FERET) Face Database. We used the same dataset used by Uzunov et al. [9]. This dataset contains 10556 gray-scale images of size 32x32 pixels face and non-face samples. There are 3156 face samples where each instance has a single sample of frontal upright human face. These images were collected from the FERET Face Database by [9], including human faces from all races with different face expressions, some wearing glasses, having beard and/or mustaches. The non-face dataset contains 7400 samples of random sampling images of size 32x32 of indoor or outdoor scenes which are collected randomly from the World Wide Web. The success rate achieved using a variable threshold on edge images is 99.9%. Our detection rate is better than the best Haar wavelet packet dictionary test results on this dataset which is reported as 99.74% with 150 atoms. Symmlet-2, Symmlet-3, and Symmlet-5 wavelet test results for 150 atoms are 99.25%, 99.51%, and 99.47%, respectively.

We also tried an AdaBoost classifier on the edge images with the same feature vectors, and achieved 94.9% success rate.

In this paper, a human face detection algorithm is proposed. A set of detailed experiments in both real-time and well-known datasets are carried out. Our experimental results indicate that support vector machines work better than AdaBoost classifiers for our feature extraction method.

REFERENCES

[1] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 23-38, Jan. 1998.

[2] K. K. Sung, and T. Poggio, "Example-based learning for view-based human face detection", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 39-51, Jan. 1998.

[3] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: an application to face detection", *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'97)*, June 17-19. 1997, pp. 130-136.

[4] A. V. Nefian and M. H. Hayes III, "Face detection and recognition using hidden markov models", *IEEE Int. Conf. on Image Processing (ICIP'98)*, vol. 1, pp. 141-145, 1998.

[5] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces", *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'91)*, pp. 586-591, 1991.

[6] E. Hjelmas and B. K. Low, "Face detection: a survey", *Computer Vision and Image Understanding 83*, pp. 236-274, 2001.

[7] C. Garcia and G. Tziritas, "Face detection using quantized skin color regions merging and wavelet packet analysis", *IEEE Trans. on Multimedia*, vol. 1, pp. 264-277, Sept. 1999.

[8] Y. Zhu, S. Schwartz, and M. Orchard, "Fast face detection using subspace discriminant wavelet features", *IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 636-642, June 13-15. 2000.

[9] V. Uzunov, A. Gotchev, K. Egiazarian, and J. Astola, "Face detection by optimal atomic decomposition", *SPIE Symposium on Optics and Photonics*, 2005.

[10] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR 2001)*, vol. 1, pp. 511-518, 2001.

[11] L. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*, Prentice-Hall, Inc., NJ, 1993.

[12] J. Kovac, P. Peer, and F. Solina, "Human skin colour clustering for face detection", *The IEEE Region 8 Computer as a tool EUROCON 2003*, vol. 2, pp. 144-148, Sept. 22-24. 2003.

[13] C. W. Kim, R. Ansari, and A. E. Cetin, "A class of linear -phase regular biorthogonal wavelets", *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 673-676, March 23-26. 1992.

[14] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, Inc., Canada, 2000.

[15] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines", 2002.

[16] CVL Face Database, P. Peer, http://www.lrv.fri.uni-lj.si/