



# Determining Effective Features for Face Detection Using a Hybrid Feature Approach

Sepideh Araban<sup>1</sup>, Fardad Farokhi<sup>2</sup>, Kaveh Kangarloo<sup>3</sup>

<sup>1,2,3</sup>Electrical Engineering Department, Islamic Azad University, Central Tehran Branch, Tehran, Iran. Email: sepide\_araban@yahoo.com

---

## Abstract

Detecting faces in cluttered backgrounds and real world has remained as an unsolved problem yet. In this paper, by using composition of some kind of independent features and one of the most common appearance based approaches, and multilayered perceptron (MLP) neural networks, not only some questions have been answered, but also the designed system achieved better performance rather than the previously presented works. The designed face detection algorithm is composed of two main stages; in the first stage of the algorithm, color based skin detection is performed using some selected color space components from the whole 15 color space components in order to reduce the search space and in the second one, verification of detected regions is done using some other kinds of different features including texture, gradient, image and geometric features. Unlike the other studied issues, in this paper, various types of features aren't evaluated in separated algorithms and systems; rather they compete all together in one competitive learning vector and after the training of the neural network the system participates in the process of feature selection. Using designed method and besides of dimensional reduction of input matrix extraordinarily, each chosen feature was ranked.

*Keywords:* face, detection, classification, feature selection, dimensional reduction, image processing

© 2012 IAUCTB-IJSEE Science. All rights reserved

---

## 1. Introduction

Face detection plays a major role in the wide range of computer vision applications such as face recognition, human computer interaction and etc. Various methods and feature sets have been used for detecting faces in images. Face detection problems includes both visible and non-visible spectrum. [1] used both visible and far infrared cameras for detecting pedestrians, in order to reduce search space. Although, methods based on non-visible spectrum such as infrared do not face many visible spectrum challenges, however, expensive equipment are necessary for these methods with tedious setup procedures which have limited their use to specific application areas such as biomedical applications [2]; furthermore, most of the applications must be performed in the visible spectrum because of

available databases. [3] performed segmentation to form a binary skin map for face detection task, by using the YCbCr color space, the Gaussian model and thresholding. In [3] the candidate regions are verified by some rules such as distances between elements of face and the existence of minimum one hole in the region. In order to overcome rotation, the candidate regions are rotated by 15 degree steps and calculations are repeated for each step. This is a very time consuming task, furthermore, due to the distortions of facial elements edges, in small faces, the skin around the elements overlap facial elements such as eyes and it reduces the existence chance of eye and lip holes in these detected regions. In general, it is impossible to translate all human knowledge into those necessary explicit rules that could be accurately comprehended by computers so that the knowledge-based methods show poor

performance when they face irregular faces or complex backgrounds that do not conformed to the determined rules [4].

In order to become invariant against various factors such as scale and to cover huge and different states in the real world, several algorithms are required for the systems based on the Template matching, knowledge-based methods and facial feature approaches. Even by using a hybrid approach, due to the fact that the issue of face detection contains huge various poses, profiles and other factors, it is impossible to cover all the possible cases using pre-defined rules and templates. Furthermore the template matching, knowledge-based methods and facial feature approaches make the system as domain specific. However, by use of learn and appearance based methods, it becomes possible to obtain the model of face including the situation of facial elements from each other more precisely, flexibly and automatically from training examples. This gives the system a generalization capability and helps it become more flexible against variation. The reliability of the system against changes is increased even more, using combination of these kinds of methods and different kinds of independent features of effective parameters, as carried out in this paper.

[5] used a set of distinctive rectangle features, called multi-block local binary patterns (MB-LBP) and developed a boosting-base learning method based on those features for detecting faces in an image. In general, features based on blocking or rectangular features, mostly suitable for window search methods because in these methods, the size of the search windows is specific and the dimensions of rectangles can be constant. In addition, the extracted features from sub images of the same objects with different sizes can be differ from the viewpoint of a system that is trained using these features. As an example, in a large image, the eyes are detected as eyes with both black and white parts but in a small face, the eyes becomes too small and may not be detectable at all. Furthermore, [5] evaluated the proposed cascade face detector on CMU+MIT face database. This database includes 130 images with 507 frontal faces. Other profiles of the face are not noted in [5].

In this work, in order to benefit from the advantages of LBP features, the LBP image moments is calculated instead of using rectangular method. In this approach, besides of considerable reduction in number of features, the features became independent from some factors such as changes in scale.

[6] used combination of the local successive mean quantization transform features and a split up Sparse Network of Winnows for face detection. They only detected frontal faces and considered that the illuminance is constant in the chosen local area. In

addition, they did not discuss how their method was made independent of the camera type. They also used a mask to remove background in the corners of an image that probably yields some important information removal particularly in the side views of faces. Furthermore, they used window search method and downscaled and resized the image with a scale factor named  $S_c = 1.2$  repeatedly.

Among the face detection methods, ones with learning based algorithms have attracted much attention and have proved their efficiency by their excellent results [7]. But the problem which is not solved yet is the face detection in highly different poses and various scales in cluttered background. Searching faces in a huge space is a difficult task with high error rate. Therefore, at first, the relative position of faces in the images is determined using independent features of many various factors to enhance the total performance and then only detected regions are verified. Due to following reasons, this method (partial search) is a better choice than window search method:

1. Due to Removal of most background textures, the volume of object and non-object distributions become closer together and therefore collecting non-object samples are facilitated.

2. Problem of complexity reduction due to omitting many background textures reduced the required patterns and inputs of the network according to PAC Learning theory [8].

3. No need to perform sweeping operations and consequently, no need to have sweeping windows and several filters in different sizes or developing pyramid images from major image in various resolutions [9, 6] in order to establish a system independent of the object size.

4. Since higher number of steps increase the operation time and cause several false detection of an object and low number of steps make the objects invisible to the detector; determining the optimum number of steps is very important and partial search method advantage is that it is no need to determine the value of the sweeping steps.

5. Unlike window search method, the process of the whole image is done homogenously.

In this paper, skin color is used as the initial independent feature. The Color processing is usually faster than other face features processing and color is independent of orientation [10]. The Color is independent of size, position, rotation and partially occlusion as well [2]. However, using only one feature type rarely allows an accurate detection, increasing the number of features could not always lead to a better result and relationships and features consistency is much more important. Recently, most features that are used to detect faces are selected from a large set of local geometric features such as

Haar wavelets [11]. Although, Haar features calculations are simple, they have a limited ability to make distinctions [11, 12]. In addition, in order to obtain an acceptable result by using Haar features, a large number of various rectangles are needed and large amount of features lead to higher calculation costs. Furthermore the size of rectangular regions is object dependent [9].

In general, there is no reason that discriminant features like Principal Component Analyze (PCA) based features perform better than the features themselves [13]. In addition, since principal components maximize the variances of the input data, so PCA contains more unwanted variations and is not robust to the independent noise of the data features [7].

[14, 15, 1, 16] have used histogram-based methods. In practice, large amount of different images have the same illumination levels or gradient orientations. In addition, in those feature type in which, only the number of mentioned parameters is counted, the position of the illumination levels or orientations from each other which contain important and distinct information of the objects are removed.

One of the most accepted and widely used computational model as a biologically inspired attention system was proposed by [17]. Computational models of visual saliency often try to mimic the behaviour of the receptive fields of ganglion cells. Because of calculations of on-center and off-center differences, expensive computations need to be done. The original work of [17] was improved in the system called VOCUS proposed by [18]. [9] reported that Visual Saliency Feature (VSF) calculates the same VOCUS features in a shorter time with no need to down sampling the original image. [9] compared features derived by: edge detectors, intensity gradients, Haar-like wavelets, VOCUS and VSFs computation separately, and reported that the best performance is achieved by VSF.

The studies show that a combination of features is more successful in object detection; however, some features not only have no effect on detection, but also have negative effect on object detection process. Therefore, in the object detection problems, using several feature sets are still required but effective features selection is necessary too. In the recently presented papers, different features consistency has been compared separately using different methods. Since experiment conditions are different in each algorithm, the evaluation becomes unreliable. In this paper; various feature types all together, in a vector are fed to an intelligent system and feature selection algorithm is used to select the more effective and consistent features.

## 2. Algorithm of Implemented Face Detection System

The Designed face detection algorithm consists of two main stages (Fig. 1), which includes a skin detector for reducing the search space and a face classifier for validation of candidate regions. At the first stage, the regions which have high probability of being face are localized base on skin color that is one of the best independent features approaches and then a binary skin map is generated. Certainly these regions include the other parts of the body skin like hands. After skin detection stage, some morphological operations are performed on the produced map as follows:

- 1- Any object smaller than 10 pixels was removed from the area of the map, because these pixels are too small to be the visible faces even by human and most of them are generated because of the noise.
- 2- Morphological closing operation on the binary map image is performed with a disk type structuring element.
- 3- Holes are filled
- 4- Morphological opening operation on the binary map image is performed with a disk type structuring element. In this stage final binary map are created.
- 5- The final binary map is labeled and bounding boxes of each region are calculated.
- 6- three kinds of cutout images are created from final binary map for each region as follows: 1- full cutout image 2- masked cutout image 3- binary cutout image

Three kinds of cut out images are generated from final binary map and gray version of the main color image which are as follows (Fig. 2): 1- Full cut out images which are cut out images of the main gray image which have the same locations and bounding boxes of the regions of final binary map. 2- Masked images which are produced by multiplying Full cut out images by the corresponding regions from final binary map. Therefore, the pixels which are labelled as non-skin ones became black (0). 3- The binary cut out images with the same regions of final binary map.

In the next stage, input feature vectors are formed by extracting the other types of feature sets from all or some of the three kinds of cutout images depend on the kind of the feature. Finally, these feature vectors are used as the input matrix of face classifier for more accurate investigation and validation of candidate regions. In fact, this method is a composition of two effective face detection approaches, independent features and appearance

based approach and generates a flexible and generalizable face detection system based on this method.

### 2.1. First Stage: Color Based Skin Detection

In this stage, the color based skin detection based on clustered based network [13] is developed as a preprocessing step for face classifier. In the training phase, after features extraction (Table.1) Each of collected skin and non-skin pixels sets are clustered into three clusters of light, middle and dark type clusters by Kmeans algorithm; the same types of the skin and non-skin clusters are fed to a separate classifier for final decision therefore three classifier are used for cluster classification. The dark, middle and light cluster based networks are all one-layer MLP with only two, three and three neurons in their hidden layers, respectively. The NN size is reduced by the method and too large NN sizes according to the image size is prevented which plays an important role in system overall speed. In the test phase and at first, feature vectors are extracted from an image then each point of input matrix is compared with the training clusters centers by Euclidian distance, and is labelled with corresponding middle, dark and light cluster type. All points are labelled together at once and maximum three new clusters are created from the input image. After clustering, each new cluster is used by the corresponding tuned network as the input matrix; therefore, each input passes through only one network and in other words the networks are not used in series. The output of this stage is a binary map as the same size as the main image and some morphological operations are performed on it. The final binary map and the grey scale image of the original image are used by the next stage. In this paper, In order to achieve higher detection rate regardless of the false alarm rate value in skin detection which is more important in this algorithm and preparing the skin detection network as a pre-processing step for a face classifier, the networks are trained more than the previous work [13].

### 2.2. Second Stage: Face Classification System

The collected face and non-face training patterns of this stage were detected from the extracted regions of previous stage from the training images. The collected patterns include faces with various sizes, different poses, profiles, orientation, types of skin, Facial expressions and even faces with glasses or hat and under different illumination condition. In addition, in order to capable the system to detect occlusion faces and images in which the total face is not detected in the first stage completely, cut out images from one fourth up to half of the face were used in the both training and test phases as well

as the full faces. After generating cut out images, other features rather than color are extracted from them and a network is trained using the complete feature matrix. The more effective face detection method is the learning based ones like neural networks (NNs) [5]. In this paper, due to the generalization ability of the MLP NN, and its ability to learn the complex non-linear input and output relationships, the MLP is used as the classifier which is one of the most common and popular type of the neural networks. The designed neural network architecture is shown in Table.2.

The Receiver Operating Characteristic (ROC) curve of classifier is plotted and two best points regarding to the classification rate and false alarm rate are chosen. After selecting the two best points of ROC curve, the feature selection algorithm is applied separately to the both networks which are related to the two points. Then, the networks are re-trained using the new selected features for verification. The new ROC curves of the new networks are plotted and the best points of each ROC curves are selected accordingly.

ROC curves were plotted as detection rate (DR) versus false alarm rate (FAR) per each epoch. It means that the network must be tested after each epoch during the train process and then the DR and FAR of the network are plotted. This process is performed on all the epochs and at last, the ROC curve is created. Finally, the best point (epoch) of this curve is chosen. This point is corresponding to the best performance of the network. Using ROC curve prevents the learn process from overtraining, determines the tradeoff between sensitivity and specificity and makes a robust comparison of the two networks.

#### 2.2.1. Features

The generated cut-out images from first stage have different sizes and their dimensions are not equal. Therefore, the generated vectors lengths of some kinds of extracted features from cut out images are different too; but the size of the vector for a MLP classifier must be equal for all patterns. In addition, the features used in the second stage must be independent of rotation, size and location in order to preserve the results of the first stage. Since most of the features have not such properties, the approach of this paper is to use the feature invariant moments instead. Due to using seven invariant moments, the feature vectors lengths become equal. These invariant moments of binary, full, masked cut out images and images which are derived from VSF, intensity gradient, edge detector and LBP are calculated.

Derived Features of VSF, edge detector and intensity gradient are not computed from masked cut out images. VSF images derived from masked cut out images were not clear because of the black parts of the masked images. The boundaries of some detected faces of the first stage were not fully fit the face area because of occlusion. These boundaries generate unreal and sharp edges because of the black parts. Therefore, in order to avoid unreal edges in the model simulation and learning process of face and non-face patterns, the edge detector and intensity gradient features are derived only from full cut out images instead of the masked ones.

Other kinds of used features are geometric features which are independent of size and are as follows:

**Solidity:** the proportion of the pixels in the convex hull that are also in the region.

**Eccentricity:** the eccentricity of the ellipse that has the same second-moments as the region. The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length.

### 2.2.1.1. Invariant geometric moments

For a digital image  $f(x, y)$ , the moment of order  $(p + q)$  per  $p, q = 0, 1, \text{ and } 2 \dots$  is defined as follows (1):

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y) \quad (1)$$

The central moments are defined as:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (2)$$

Where,

$$\bar{x} = \frac{m_{10}}{m_{00}}, \quad \bar{y} = \frac{m_{01}}{m_{00}} \quad (3)$$

The normalized central moments, computed as follows:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad (4)$$

Where:

$$\gamma = \frac{p+q}{2} + 1 \quad (5)$$

And finally, the seven invariant moments are obtained from second and third moments:

$$\phi_1 = \eta_{20} + \eta_{02} \quad (6)$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \quad (7)$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \quad (8)$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad (9)$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (10)$$

$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (11)$$

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (12)$$

This set of moments is invariant to the transfer, rotation and scale ([19]).

### 2.2.1.2. Local binary patterns moments (LBPM) features

It has been proved that LBP is a powerful feature set in texture classification [20, 21] and this texture model is independent to any grey scale homogenous conversion [16]. The calculation is as follows:

The pixels of each  $3 * 3$  neighborhood are labeled by amount of the central pixel.

$$s(g_0, g_i) = \begin{cases} 1, & g_i \geq g_0, \\ 0 & g_i < g_0, \end{cases} \quad 1 \leq i \leq 8, \quad (13)$$

The eight differentiation signs are coded to an 8-bit digit in order to obtain the amount of central pixel.

$$LBP(g_0) = \sum_{i=1}^8 s(g_0, g_i) 2^{i-1} \quad (14)$$

Finally, seven invariant moments are computed from the image which is achieved by computing LBP from the original image.

Since the rectangular methods such as MB-LBP [5] are compatible with window search methods therefore, in this paper, in order to gain advantages of LBP features besides of considerable reduction in the number of features, the moment of LBP Image was computed instead of using rectangular method.

Furthermore, by using this method, features which are an image themselves, become independent of several factors such as scale variation.

### 2.2.1.3. Visual saliency moments feature (VSMF)

In this work, the proposed VSMFs are a modified version of the same features that were used in [9]. The high computations of applying attention systems on full resolution image have prevented the direct use of saliency mechanisms as a feature extraction method. The VSFs [9] use the same biological theory as [17, 22] but it uses an integral image on the original scale in order to obtain high quality features in real time.

In this work, the seven invariant moments of VSFs highly reduced the dimension of the feature vectors as well. This new feature is called VSMF.

The VSMFs are computed as follows:

At first, surround and center are defined using (15).

$$\begin{aligned} \text{surround}(x, y, \zeta) &= \frac{\text{recSum}(x - \zeta, y - \zeta, x + \zeta, y + \zeta) - i(x, y)}{(2\zeta + 1)^2 - 1} \\ \text{center}(x, y) &= i(x, y) \end{aligned} \quad (15)$$

In the second, each pixel of each intensity sub maps is calculated using (16)

$$\begin{aligned} \text{Int}_{on,\zeta}(x, y) &= \max\{\text{center}(x, y) - \text{surround}(x, y, \zeta), 0\} \\ \text{Int}_{off,\zeta}(x, y) &= \max\{\text{surround}(x, y, \zeta) - \text{center}(x, y), 0\} \end{aligned} \quad (16)$$

Where on and off represent the on-center and off-center differences respectively and  $\zeta \in \{1, 2, 4, 8, 16, 32, 64, 128\}$  is the surround.

Then, an on-center intensity map is calculated by summing the six on-center intensity sub maps by pixel and the off-center intensity map is produced in the same manner. Both maps are summed finally.

$$\begin{aligned} \text{Int}_{on} &= \sum_{\zeta} \text{Int}_{on,\zeta} \\ \text{Int}_{off} &= \sum_{\zeta} \text{Int}_{off,\zeta} \end{aligned} \quad (17)$$

In the final stage, the seven invariant moments are calculated from the image resulted from the sum of the two maps.

### 2.2.1.4. Edge detection

Spots in the image where the intensity changes rapidly are considered as an image edges and determined using the two criteria as follows:

- Places where the first derivative of the intensity is larger in magnitude than some threshold.
- Places where the second derivative of the intensity has a zero crossing.

The most powerful edge detection method which is reported previously is the canny method, and therefore is used in this paper. The difference between canny methods and other edge detectors lies in detecting both weak and strong edges using two different thresholds and only the weak edges which are connected to the strong edges are included in the output; this method is therefore more stable against noise, and more likely to detect true weak edges. Since face details are potentially important in detection, the canny method is used together with the seven invariant moments of the edges as another feature.

### 2.2.2. Dimensional reduction and feature selection

Feature vector dimension is reduced considerably using the seven invariant moments calculation of the feature sets which are an image themselves. As an example, extracted feature vector from a 30\*30 cut out image has 7202 elements; the length of feature vectors decrease to 58 and become equal to each other for all cut out images of different sizes, using the moments calculation; while most of the cut out images are larger than 30\*30 and also the number of features become even less than 58 by feature selection. In this paper, the Utans algorithm is applied as the feature selection method; it is reported as the best method in [23]. In order to compare the system performance before and after feature selection, the ROC curves are used.

## 3. Experiment Results

The presented algorithm was implemented by MATLAB. 883 cut out images of human faces and 832 non-face cut out images were collected for the face classifier system. Those images were selected from the detected regions of 2501 images which were the output of the skin detection stage. The 2501 other images were used for total face detection system performance evaluation. Those 5002 Images are extracted from VOC2007 database. The VOC2007 database is a relatively new, highly complicated and contains various different Images. The VOC2007 database contains interpreted images of real world with highly cluttered background in various textures under different light sources and illumination, orientations (even reversed), profiles, poses, facial expressions and for 20 different types of objects. Of course, this database has been designed for detecting person and contains faces in any ages,

positions and sizes. It includes different types of human pedigrees and occlusion (a person wears glass and hat and running a bicycle); moreover, there is no person in some of its images at all. This database was initially designed for competition purposes. A number of patterns used in the second stage are shown in Table.2.

As it was explained before, in this paper, In order to find desired points for face detection task, the previous skin classification system [13] is trained more using the selected features to obtain higher DR and so more points of ROC curves of skin classifiers are plotted. The ROC curves of the new clustered based networks and the selected points are shown in the Fig.3 and the Table.3 respectively. But Due to no necessity of more training in dark cluster case because of the 100% DR, the training was not continued more. After applying the new trained skin classifier system on all training images, cut out images are collected from detected regions. Then, in the second stage, features are extracted from training cut out images and the face classification system is trained using this new feature vectors. In order to select the optimized points, ROC curve of the face classifier is plotted, as shown in Fig.4.a. The selected points are listed in Table.4.

In the next section, feature selection algorithm was applied to both trained face classification networks per both selected points and the ROC curves of both cases were plotted (Fig.4.b and Fig.4.c).

The results of this section show slight differences with the results obtained from classifier system that have been trained and tested by using whole features (Table.4 and Fig.5). The results of applying the feature selection algorithm are listed in Table.5. The error is the difference between number of errors before and after performing feature selection algorithm. The total error equals the sum of seven errors related to seven invariant moments.

#### 4. Discussion

As a result of using color features in the first stage, the system has become independent of size, location, angle, rotation and Face regions of different sizes and locations has been localized. Therefore the next chosen feature sets also must be independent of these parameters for preserving this independence property. Furthermore, the input vector dimension of a MLP neural network must be the same in size for all patterns but the detected regions sizes of the first stage are different therefore the number of extracted features from those become unequal for the most of feature sets and types. Furthermore, most of the feature sets have not independence property. In this paper, each dependent feature set which is itself an image becomes

independent by using another feature type which is computed for total pixels in an image. This features type set is consist of seven invariant geometric moments which are independent of size, rotation, orientation. Another reason for using seven invariant geometric moments of feature sets instead of using themselves directly is size reduction of input feature vector. Using discussed feature sets directly causes extraordinary high feature vector dimensions. As a result of [24], the high dimensions of input matrix reduce the performance of a MLP Neural network dramatically. They compared row pixels and pixel statistic as a feature vectors and performed face detection using a MLP. Then they reported that high input dimensions lead to low neural network performance and so they used mean and standard deviation instead of row pixels as features.

In this paper, in order to select the effective features and solve the mentioned problem, seven invariant geometric moments of feature sets were used and a hybrid feature vector was constructed. The different features have equal chances to be selected by the classifier and finally 30 and 23 features were selected among 58 reduced features for the two mentioned classifier.

As it was explained before, two kinds of gray cut out images and one kind of binary image were generated per each region of the first stage images.

Determined effective features of both selected points from ROC curve were similar to each other. The results are shown in Table.5 and VI and they were as follows. The invariant moments of detected edges from full cut out images, had the most positive effect on face detection performance. The next feature set was the moments of row pixels from full and masked cut out images respectively and had close values to each other. LBPMs of masked images feature set were at the third degree of effectiveness. Feature sets in the fourth degrees were the moments of binary image and intensity gradient of cut out images. Feature sets in the next degrees were VSMF and LBPM of full cut out images. Eccentricity hadn't any positive effect in classification and even was determined as an undesired feature. The Solidity feature had negative effect on the one point and a few desired effects on the other one so the geometric features were not evaluated as desired features and were the last features in the ranking. Obviously the first and second invariant moments were the most effective ones among invariant moments and the other invariant moments were required for obtaining a desired performance too. Only, LBPM features which were computed from masked cut out images, was determined as more effective features than those which used the full ones.

It was shown that the most desired feature was the moment of detected edges image computed from

full cut out images. One reason is that edges image has less redundant information in respect to the other features and contains the main information of facial features. A network are able to obtain more accurate model of an object When it learn patterns of that object edges because it do not face a huge set of complex and semi random information. Furthermore, this model can contain information such as form, structure and statements of facial features which are trained automatically and for various poses, orientation and etc. So it has generalization ability and do not have disadvantage of Template matching and knowledge-based methods. The reason that gradient of an image which acts like a high pass filter, did not show an expected efficiency, can be returned to edge detection algorithm. The canny method is less like the other edge detectors to be fooled by noise and more likely to detect true weak edges due to detecting strong edges and week edges which are connected to the strong edges. Facial details are important for face detection but detecting edges in a wide range, such as gradient process, increase redundant information that decrease simulated models accuracy. Geometric features are useful when the detected faces regions are in general and common positions and also they do not have any occlusion which destroys the elliptical boundaries of faces. Whereas, in this paper, many of cut out images are a half or a quarter of face so their binary patterns have not an elliptical shape and they can be similar to binary patterns of many other objects. VSF features did not show facial and edges details strongly because of multi summation process. Due to multi summation or multiple low pass filters in VSF computation process, the difference between the values of boundary pixels and their surrounded ones decreases and therefore, distinguishing main facial features from the background gets difficult because of weak edges and boundaries. The experimental result was clearly in accordance. Removing VSF features from the input vector did not increase the amount of the error very much.

Certainly, still shadows in the image can produce unreal and strong edges in the image. Also edges of gradient image and image of detected edges are stronger than image itself due to the edge detection process and gradient computation of the image. Therefore edge and gradient features could not make row pixels features out of the competition completely Because of those unwanted sharp edges and their more undesired influence. Another reason is week edges of small scale face images. These images are blurred and their facial details do not produce strong edges. These images can be ignored by the edge detection based face detection systems and so row pixels features are still required for a proper performance. However edge features and so high

frequency information of an image have more desired influence than the other features but effective composition of features are still required for a proper face detection system performance and make the system more robust to cluttered real word with huge and various kind of data.

Experimental results showed that LBP of masked cut out images have better performance than the full cut out ones. It has been shown that in this type of feature, negative effect of redundant marginal is greater than the negative effect of some of the desired objects parts from an image. Finally Table.7 shows comparison between designed method and the previously presented methods on the same database.

## 5. Conclusion

In this paper, a face detecting system in color and cluttered real word images which was taken in visible light (VIS) spectrum was presented. For this purpose, composition of some kind of independent features and one of the most common appearance based approaches, multi-layered perceptron (MLP) neural network was used. In the first stage, the color based skin detector was developed to reduce the search space. In the second one, a face classification system was designed for verification and refusal of localized regions derived from the first step. Moment computation of some feature sets was performed for realizing three purposes: 1- input matrix dimensionality reduction 2- feature vectors dimensionality equalization 3- to make features independent. The designed system using one single-hidden layer MLP with three hidden neurons and one clustered based MLP and showed better performance than other investigated previous works. As the results of applying the Utans algorithm, it was shown that the moments of detected edges of the full cut out images, was determined as the most effective features and the moments of row pixels of the full cut out images, determined as the second in features rank. The solidity feature and Eccentricity feature was evaluated as the weakest features. VSMFs and LBPMs features were penultimate in the ranking. High frequency information of image had the high degree of desired influence and major effect on face detection. In general, Gradient kinds of features had more desired influence than the image ones. But certainly effective composition of features is still required for a proper face detection system performance and makes the system more robust to huge and various kinds of data in the cluttered real word images. Each of masked, full or binary cut out images could be useful and it depends on the type of the feature. Since removing redundant marginal could not be accurate and some important part of the object in the image would be removed in some cases, so it is better to choose a full image and in the cases



which removing redundant marginal has more desired effects, masked images are better to be chosen instead and it could be identified by feature selection. The first and the second invariant moments had more positive effects than the other ones but in order to achieve the proper performance, computing and using all invariant moments was necessary in some features sets.

## References

- [1] M. Bertozzi, A. Broggi, M. Del Rose, M. Felisa, A. Rakotomamonjy and F. Suard, "A Pedestrian Detector Using Histograms of Oriented Gradients and a Support Vector Machine Classifier", In Proc. IEEE Intelligent Transportation Systems Conference, pp.143-148, 2007.
- [2] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A Survey of Skin-Color Modeling and Detection Methods", Elsevier Pattern Recognition., Vol.40, No.3, pp.1106 – 1122, 2007.
- [3] Wen-Hsiang Lai, Chang-Tsun Li, "Skin Colour Based Face Detection in Color Images", IEEE International Conference on AVSS06, IEEE Computer Society, pp.56-61, 2006.
- [4] Chang Huang, Haizhou Ai, Yuan Li, and Shihong Lao, "High-Performance Rotation Invariant Multiview Face Detection", IEEE Trans. on Pattern Analysis and machine intelligence ACTIONS, Vol.29, No.4, pp.671-686, 2007.
- [5] Lun Zhang, Rufeng Chu, Shiming Xiang, Shengeai Liao, and Stan Z. Li, "Face Detection Based on Multi-Block LBP Representation", Center for Biometrics and Security Research & Natuinal Laboratory of Pattern Recognition, China, 2007.
- [6] M. Nilson, J. Nordberg, and I. Claesson, "Face Detection Using Local SMQT Features and Split up snow Classifier", IEEE ICASSP'07, Vol.2, pp.589-592., 2007.
- [7] Ming-Hsuan Yang, David J. Kriegman, and Narendra Ahuja, "Detecting Faces in Images: A Survey", IEEE Trans. On Pattern Analysis and Machine Intelligence., Vol.24, No.1, pp.34-58, 2002.
- [8] L.G. Valiant, "A Theory of the Learnable", Communications of the ACM, Vol.27, No.11, pp.1134–1142, 1984.
- [9] S. Montabone, and A. Soto, "Human Detection Using a Mobile Platform and Novel Features Derived from a Visual Saliency Mechanism", Image and Vision Computing., Vol.28, No.3, pp.391-402, 2010.
- [10] Sanjay Kr. Singh, D.S. Chauhan, Mayank Vatsa, Richa Singh, "A Robust Skin Color Based Face Detection Algorithm, Tamkang Journal of Science and Engineering, Vol.6, No.4, pp.227-234, 2003.
- [11] Paul Viola, Michael Jones, "Robust Real-Time Object Detection", International Journal of Computer Vision (IJCV), Vol.2, pp.137–154, 2004.
- [12] S.Z. Li, ZhenQiu Zhang, "Float boost Learning and Statistical Face Detection", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.26, No.9, pp.1112–1123, 2004.
- [13] S. Araban, F. Farokhi, K. Kangarloo, "Determining Effective Colour Components for Skin Detection Using a Clustered Neural Network", IEEE International Conference Signal and Image Processing Applications (ICSIPA11), pp.541-546, 2011.
- [14] L. Zhang and R. Nevatia, "Efficient Scan-Window Based Object Detection Using GPGPU", IEEE Computer Society Conference on Computer Vision and Pattern Recognition workshops (CVPRW08), pp.1-7, 2008.
- [15] N. Dalal and B.I Triggs, "Histograms of Oriented Gradients for Human Detection", IEEE Computer Society Conference on Computer Vision and Pattern Recognition workshops (CVPR05), Vol.1, pp.886-893, 2005.
- [16] H. Zhang, W. Gao, Xilin Chen, Debin Zhao, "Object Detection Using Spatial Histogram Features", Image and Vision Computing, Vol.24, No.4, pp.327–341, 2006.
- [17] L. Itti, "The ilab Neuromorphic Vision C++ Toolkit: Free Tools for the Next Generation of Vision Algorithms", The Neuromorphic Engineer, Vol.1, No.1, 2004.
- [18] S. Frintrop, "VOCUS: A Visual Attention System for Object Detection and Goal directed Search", PhD thesis, Rheinische Friedrich-Wilhelms University, Bonn Germany, 2005.
- [19] R. C. Gonzalez, R. E. Woods, "Digital Image Processing", Prentice Hall Edition, 2008.
- [20] M. Pietika'inen, T. Ojala, Z. Xu, "Rotation-invariant Texture Classification Using Feature Distributions", Pattern Recognition Vol.33, pp.43–52, 2000.
- [21] T. Ojala, M. Pietika'inen, T. Ma'enna', "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.24, No.7, pp.971–987, 2002.
- [22] L. Itti, C. Koch, E. Niebur, "A Model of Saliency-based Visual Attention for Rapid Scene Analysis", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.20, No.11, pp.1254–1259, 1998.
- [23] M. F. Rodondo, and C. H. Espinosa, "A Comparison Among Feature Selection Methods Base on Trained Networks", IEEE NNSP'99 Workshop, pp.205-214, 1999.
- [24] B. A. Chin, "Object Detection using Neural Networks and Genetic Programming", PhD thesis, Victoria University Of Wellington, School of Mathematics, Statistics and Computer Science, 12 Oct., 2007.

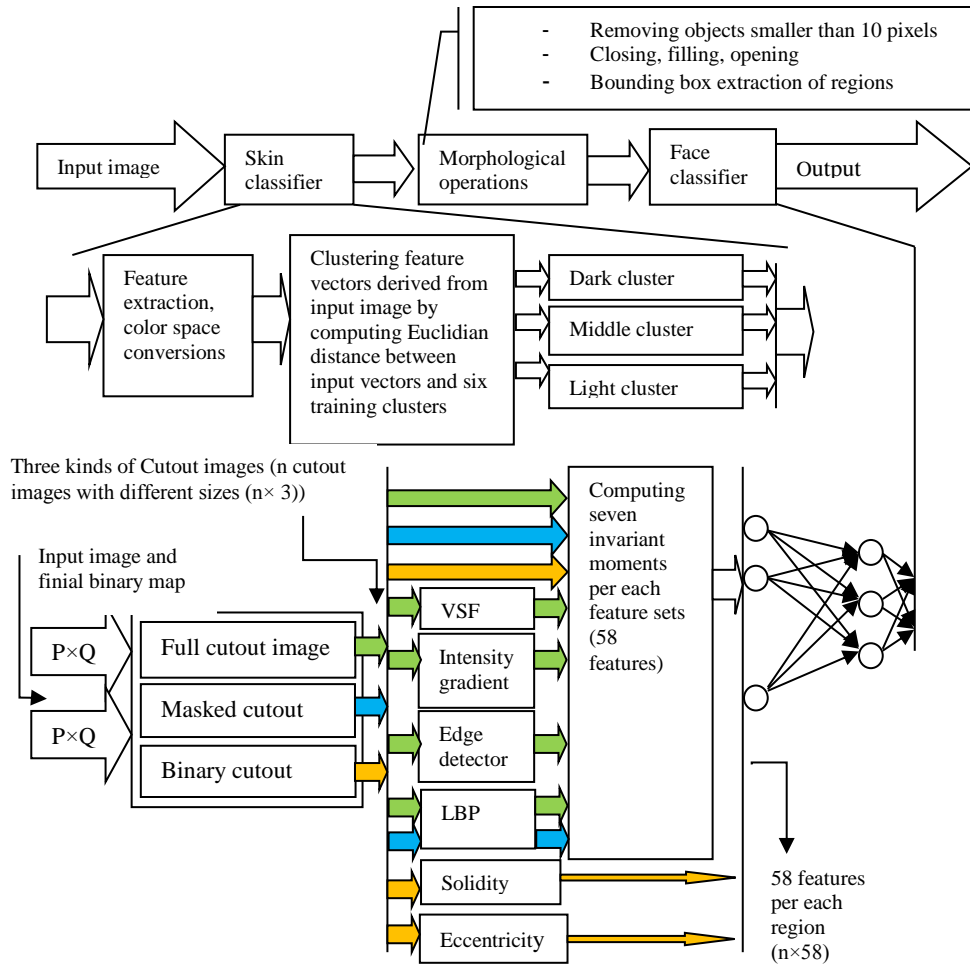


Fig.1. The block diagram of the presented face detection algorithm

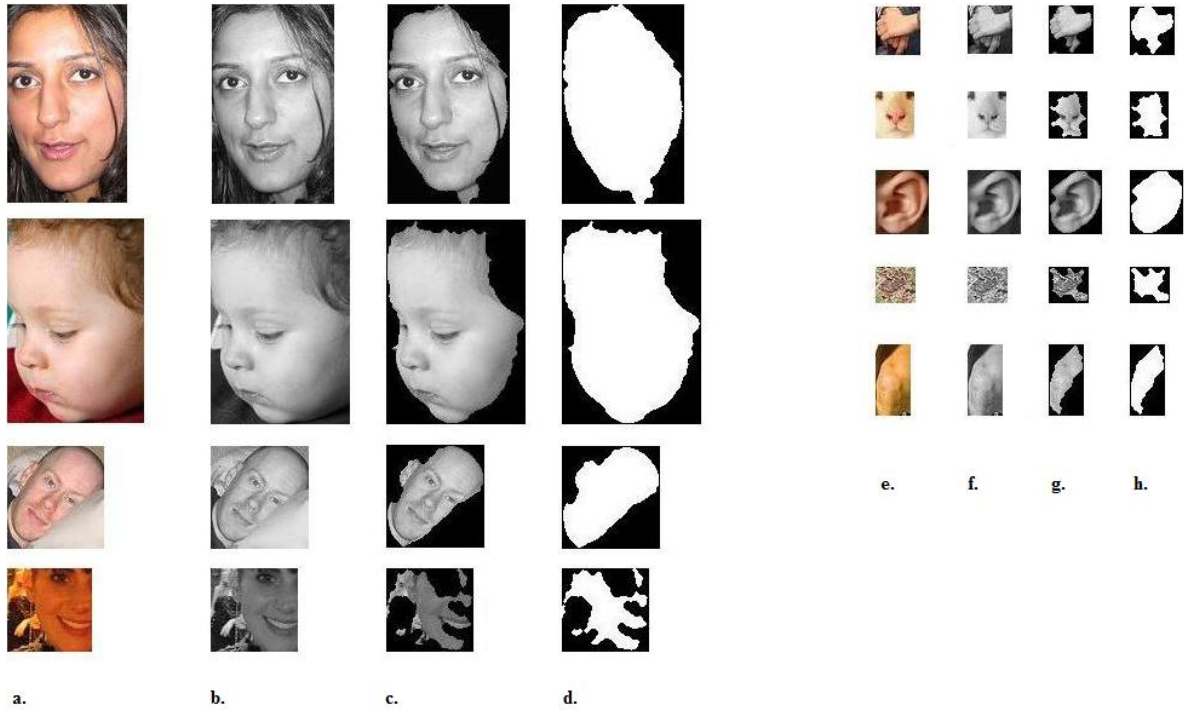


Fig.2. Three different kinds of cut-out images which are generated from the gray version of the main color image and final detected binary map of the first stage of the algorithm. On the left, There are some used cut-out faces and On the right, There are some used cut-out images of cluttered backgrounds and three kinds of corresponding generated cut-out images: a, e) the main cut-out images, b, f) The full cut-out images which are cut-out images of the main gray image which have the same locations and bounding boxes of the regions of final binary map, c, g) The masked images which are produced by multiplying Full cut-out image by the corresponding regions from final binary map. Therefore, the pixels which are labelled as non-skin ones became black (0), d, h) The binary cut-out images with the same regions of final binary map.

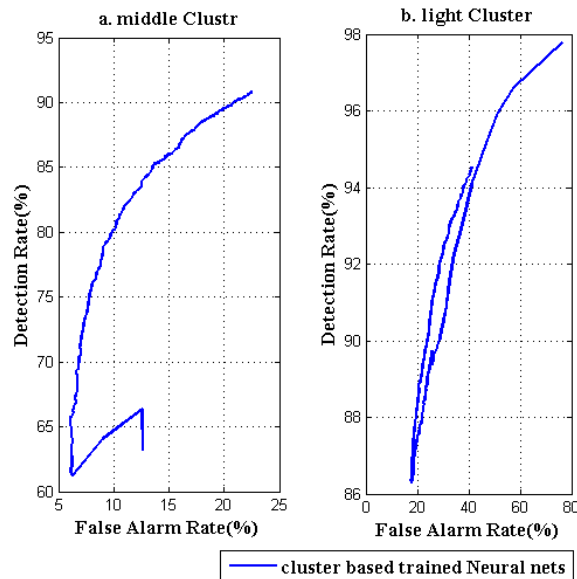


Fig.3. ROC curves which are DR vs. FAR per epoch (variation of DR vs. FAR of each training iteration) for cluster based networks a) middle cluster b) light cluster.

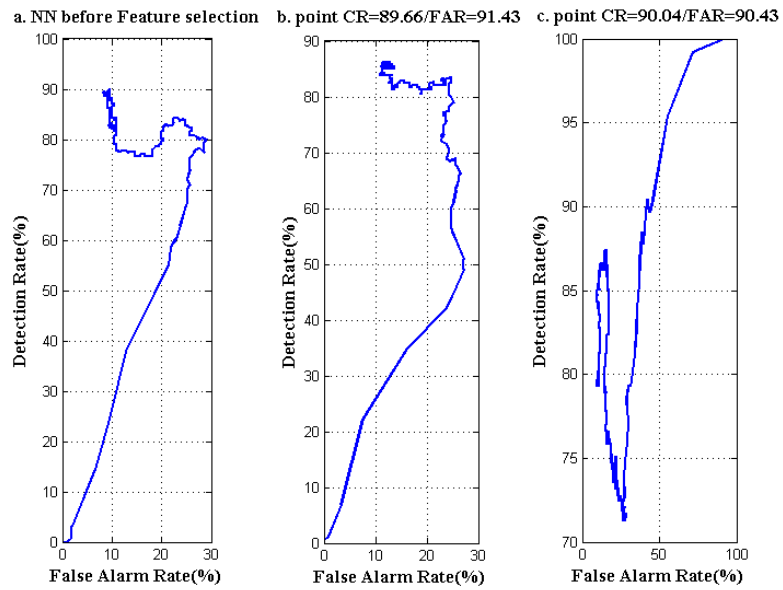


Fig.4. ROC curves of the face classifier a) 58 features. Before applying feature selection algorithm b, c) after applying feature selection algorithm to the both trained networks related to two selected points of a.

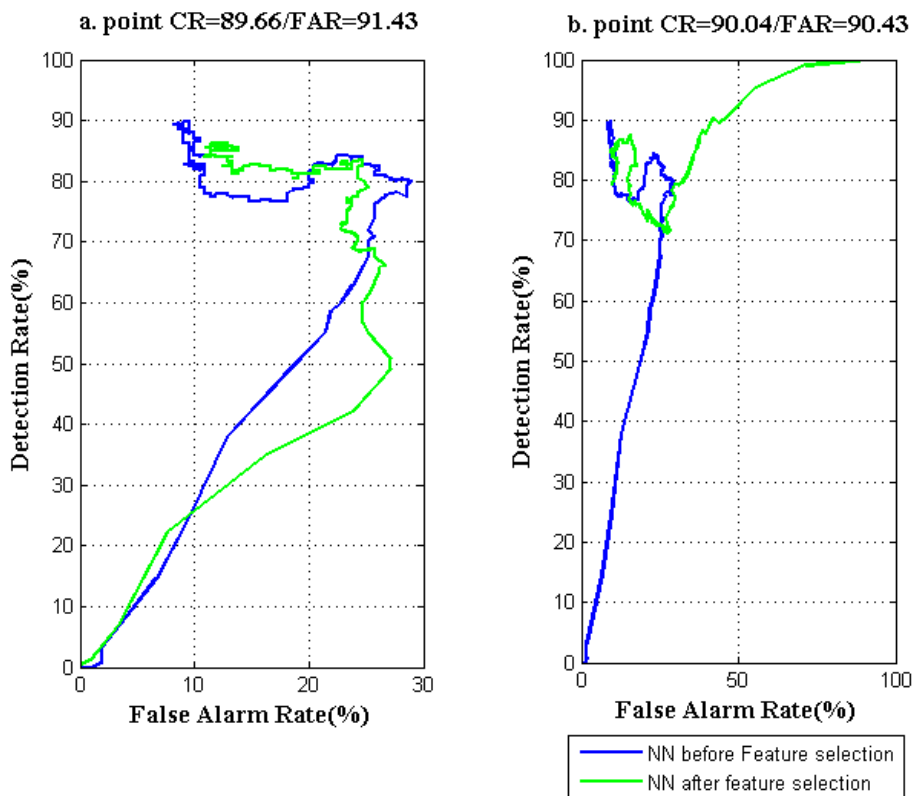


Fig.5. The comparison between ROC curves of the trained networks before and after feature selection. a) after applying feature selection algorithm to trained neural network related to the point CR=89.66, FAR=91.43, b) after applying feature selection algorithm to trained neural network related to the point CR=90.04, FAR=90.43 .

Table.1.

All of Used Color Spaces Components and Their Related Color Spaces are Presented in each Rows, Feature Selection Was Done by Utans Algorithm on Dark (C3), Middle (C1) and light (C2) clusters related NNs. (the used Color Components are Earmarked with Gray colors. sum means combinations of ratios-R/B+R/G+B/G)

All of color components of used color spaces															
C1	RGB	rgb	R/BR/GB/G	sum	YIQ	YcbCr	HSV	YcgCr	XYZ	L*a*b*	CMYK	u/v/L	xyY	L*ch	uvL
C2	RGB	rgb	R/BR/GB/G	sum	YIQ	YcbCr	HSV	YcgCr	XYZ	L*a*b*	CMYK	u/v/L	xyY	L*ch	uvL
C3	RGB	rgb	R/BR/GB/G	sum	YIQ	YcbCr	HSV	YcgCr	XYZ	L*a*b*	CMYK	u/v/L	xyY	L*ch	uvL

Table.2.

Characteristic of Applied NN Structure for face detection before applying Feature Selection which shows the Numbers of train and test Patterns of designed MLP network.

Characteristic of NNs	Multi-Layer Perceptron (MLP)					
	Algorithm: feed forward ·back-propagation· gradient descent					
	Number of layers	Units in input/hid - den/output layers	Random range	Learning rate	Train examples	Test examples
Face class	1	58-3-2	[-.5,.5]	.01	622	261
Non face class					622	210

Table.3.

The selected points of ROC Curves –detection rate vs. false alarm rate per epoch- of clustered based neural networks in order to obtain higher detection rate for using skin classifier NNs as a pre-processing for face classifier system

Cluster	Sensitivity (%)	Specificity (%)
Light Cluster	88.26	80.03
Middle Cluster	90.10	78.98
Dark Cluster	100	100

Table.4.

Number Of Input Neurons After Applying Feature Selection Algorithm (Utans) on two Selected Points of ROC Curves of Designed Neural Network for Face Classifier and the performance Results

Face classifier networks		Before feature selection (Fig.4.a) ↓				After feature selection (Fig.4.b, 4.c) ↓						
Two selected points of Fig.4a curve ↓	Class ↓	Number of neurons	Sensitivity (%)	Specificity (%)	Accuracy (%)	Description	Class	Number of neurons	Sensitivity (%)	Specificity (%)	Accuracy (%)	
												Point 1 →
Non face	91.43	89.66	90.43	Non face	88.57	86.21	87.26					
Point 2 →	Face	58	90.04	90.43	90.23	One selected point of Fig. 4.c curve. →	Face	23	86.21	89.05	87.47	
							Non face		89.05	86.21	87.47	
	Non face		90.43	90.04	90.23		The other selected point of Fig.4.c curve. →	Face	23	86.59	88.10	87.26
								Non face		88.10	86.59	87.26

Table.5.

The Result of Applying Utans Algorithm -Feature Selection Algorithm - to two Networks Related to two Point of the ROC Curve for all Kinds of Used Feature Sets. The error is the difference between number of errors before and after performing feature selection algorithm. The total error equals the sum of seven errors related to seven invariant moments.

Feature type	Cut-out image type	Point (Network)	Error							Total error
			Seven invariant moments ↓							
			$\phi_1$	$\phi_2$	$\phi_3$	$\phi_4$	$\phi_5$	$\phi_6$	$\phi_7$	
Row pixels moments	Full	1	-64	-62	-3	0	-3	4	0	-128
		2	-62	-64	2	1	-10	5	0	-128
	Masked	1	-99	-2	-1	0	0	0	0	-102
		2	-101	1	-1	0	0	0	0	-101
	Binary	1	-22	-14	-1	0	0	0	0	-37
		2	-34	-17	-1	0	0	0	0	-52
LBPMs	Full	1	-1	-1	-1	0	0	0	0	-3
		2	-5	-1	-1	0	0	0	0	-7
	Masked	1	-98	0	0	0	0	0	0	-98
		2	-92	0	0	0	0	0	0	-92
VSFMs	Full	1	-6	-5	0	-2	0	0	-2	-15
		2	-7	-4	1	2	0	0	-2	-10
Edge moments	Full	1	-71	-106	-1	-9	0	-2	-9	-198
		2	-62	-94	1	-10	0	2	-10	-173
Gradient moments	Full	1	-2	-11	-4	-21	-1	0	-5	-44
		2	0	-12	2	-26	0	1	-10	-45
Geometric features	Cut-out image type	Point (Network)	Error							Total error
Solidity	Binary	1	1							1
		2	-4							-4
Eccentricity	Binary	1	2							2
		2	3							3

Table.6.

Determined effective features of both selected points from ROC curve were similar to each other and were as follows. The invariant moments of detected edges from full cut-out images, had the most positive effect on face detection performance and the geometric features were not evaluated as desired features and were the last features in the ranking.

Type of feature	Edge moments	Row pixels moments	Row pixels moments	LBPMs	Row pixels moments	Gradient	VSFMs	LBPMs	Solidity	Eccentricity
Type of cut-out image	Full	Full	Masked	Masked	Binary	Full	Full	Full	Binary	Binary
30 Effective features (Point1)	$\phi_2, \phi_1, \phi_4, \phi_7, \phi_6, \phi_3$	$\phi_1, \phi_2, \phi_3, \phi_5$	$\phi_1, \phi_2, \phi_3$	$\phi_1$	$\phi_1, \phi_2, \phi_3$	$\phi_4, \phi_2, \phi_7, \phi_3, \phi_1, \phi_5$	$\phi_1, \phi_2, \phi_4, \phi_7$	$\phi_1, \phi_2, \phi_3$	No	No
23 Effective features (Point2)	$\phi_2, \phi_1, \phi_4, \phi_7$	$\phi_2, \phi_1, \phi_5$	$\phi_1, \phi_3$	$\phi_1$	$\phi_1, \phi_2, \phi_3$	$\phi_4, \phi_2, \phi_7$	$\phi_1, \phi_2, \phi_7$	$\phi_1, \phi_2, \phi_3$	Yes	No
Gain	1	2	3	4	5	6	7	8	9	10

Table.7.  
Comparison between The presented method and [24] which was evaluated on the same database

Method	Features	Number of training examples	Number of test examples	Cut-out images	Number of features	Classifier	Database	Evaluation
Presented method	Composition of features Table.6	1244	471	Cluttered background Frontal, profile, and other various orientations approximately [-100,100] and different rotations [-90, 90] occlusion, a quarter, half, central of faces and full faces,----	58	MLP with three hidden layer neurons	VOC 2007	ACC= 90.43% Fig.4, 5 Table.4.
					23			ACC= 87.47% Fig.4, 5 Table.4
Barret A.Chin ,Mengjie Zhang [24]	Standard deviation	150	150	---	Maximum 9×12	Genetic programming	VOC 2007	ACC= 56.60%
Barret A.Chin ,Mengjie Zhang [24]	Standard deviation	150	150	---	12	MLP with eight hidden layer neurons	VOC 2007	ACC= 66.42%