Face Detection Based on Skin Color Modeling and Modified Hausdorff Distance

Khalid Mohamed Alajel, Wei Xiang, and John Leis
Faculty of Engineering and Surveying
University of Southern Queensland
Toowoomba, QLD 4350, Australia
E-mail: {*khalid.alajel, wei.xiang, john.leis*}*@usq.edu.au*

***Abstract*- This paper presents a new face detection approach which is capable of detecting human faces from complex backgrounds. A new skin color modeling process is applied to the face segmentation process. Image enhancement is then used to improve the features of face candidates before feeding to the face object classifier which is based on a modified Hausdorff distance. The overall performance of the face detection system is evaluated and achieved a success rate of 87.5 %.**

I. INTRODUCTION

Face detection is a necessary first step in face recognition systems with the purpose of localizing and extracting the face region from the background. The human face is a dynamic object and displays a high degree of variability in appearance. In real life situations, different illumination, distance from the imaging device, occlusion, and rotation of the head in different planes are bound to happen, which poses a significant challenge to face detection. Most systems assume a specific orientation of the face such as frontal or near frontal face orientation to simplify the problem [1]-[3].

Face detection techniques can be classified into two schools: (i) image based techniques and (ii) feature based techniques. Image based techniques address face detection as a general recognition problem whereby pattern recognition is applied to the whole image. The face knowledge is therefore implicit to the users and mapping and training schemes are utilized to achieve what is known as recognition. Feature-based techniques build in explicit knowledge where features representing face as defined by the designer are first extracted from images [4]. Face detection is thus achieved by verifying that a certain degree of confidence that features extracted from images represent a face.

Color analysis on images has long been used as a technique which can give additional dimension to image compared to grey scale image. Classification is easier to handle in color space compared to gray scale. It is a known fact that the skin color of different races tend to cluster in close proximity in the normalized color space or chromaticity space. Using this skin color model, a skin candidate region is identified based on certain threshold value. Image matching methods can be well used to find correspondence between a template and given portion of an image having the most partial similarity [5].

Skin color has long been used for detecting skin color region and even in head detection system for searching head region. The major problem with using skin color model however is that it is subject to variation in illumination and hence not robust enough in detecting head candidate. One more prominent problem is that most of the time, object which appears to have skin color is not necessarily the human face, worse, it may not even be part of the human skin. It is therefore impossible to rely solely on skin color alone as an effective face detection strategy.

The use of variants of the Hausdorff distance has recently become more and more popular in image matching applications [5]. It has the advantage of being scale invariant, illumination invariant and robust in complex background. The Hausdorff distance is a robust technique used in image matching problem [6]. Traditionally, it has always been used in gray scale image to locate image candidate which is a closest match to an object. In order to search for the possible image candidate, the system generally needs to scan through the whole image until it reaches the targeted candidate. As the image size grows, so does the computing power needed to locate the image candidate. It is therefore believed that using skin color filter along with Hausdorff distance will target the shortcomings of both these strategies. Skin color filter will help identify the image candidate so that Hausdorff distance will be able conserve computing power on the image candidate while Hausdorff distance will verify the validity of the image candidate which is not possible by using skin color filter alone.

In this paper, we present a framework based on feature-based approaches where low level features are derived prior to knowledge-based analysis. In this instance, color information is used as the low level features to differentiate between face candidate and non face candidate where face candidate comprises human skin color pixels. A second low level feature which is the edges of the face candidate region and any facial features it may contain are then used. The last step of this approach is generalized measure analysis where distance between face candidate region and face template is calculated to decide if the face candidate region is really a face. The classical Hausdorff distance is simple, but it is sensitive to degradation, such as noise and occlusions. Therefore, a modified Hausdorff distance which is more robust to degradation is used instead. Modified Hausdorff distance [7] between the edges of face candidate region and that of the face template is calculated. The face candidate region is then considered a face if the modified Hausdorff distance measure is smaller than a threshold value predetermined.

This paper is organized as follows: Section II presents the proposed face detection system and results are presented in Section III. Finally, concluding remarks are drawn in Section IV.

II. SYSTEM DESCRIPTION

The overall face detection system consists of three stages: (i) Image acquisition; (ii) Skin color segmentation; and (iii) modified Hausdorff object classification as depicted in Fig 1. The knowledge base of the system acts as a supporting module which assists the various major modules in the thresholding process. The system starts with the image acquisition module which takes in color images in various formats supported by MATLAB such as BITMAP, JPEG and TIFF. In this experiment, the image acquisition process is simplified into feeding downloaded color images into the system. The color images which may contain face or non face objects are then processed by the skin color segmentation module and modified Hausdorff object classification module, respectively.

A. Skin Color Segmentation model

The main objective of skin color segmentation is to segment out any object which resembles the human skin color within a certain degree of confidence as determined by the thresholding process. Regions with low probability of being the face region will be first eliminated so that the computation power can be focused solely on regions with higher probability. The Enhancement process is then needed as there are always possibilities that some face region pixels which are discarded or non face pixels which are retained. This process thus help reconstructs the face region should part of the face pixels is unintentionally discarded and also eliminate regions which are not face and yet unintentionally retained.

1) Skin Color Filtering: The objective of skin color filtering is to perform color segmentation on the input image based on human skin color in a suitable color space. This work chooses YCbCr color space which is more tolerant towards lighting condition compared to RGB color space [8], [9]. The skin color filter is assembled based on adequately large sample of different races in YCbCr color space. In this experiment, a total of 200 face images across different ethnic groups and gender from the Aleix face database [10] are used to construct the skin color filter.

The probability of an image pixel belonging to the human skin color can be computed using the following formula as proposed by Liu [1].

$$Likelihood = p(r,b) = \exp[-0.5(x-m)^T C^{-1}(x-m)] \quad (1)$$

where $x = (r,b)^T$, x is the Cr and Cb value of individual pixel in matrix form, C is the covariance of x in the matrix form, m is the mean of Cr and Cb value in matrix form and T is a transpose of x. Using equation 1, the probability of each input image pixel belonging to the human skin falls between the range 0 to 1 where 1 is the highest possibility while 0 means that the particular pixel could never be part of the human skin.

2) Enhancement process: The basic idea of the enhancement phase is to group skin pixels in close vicinity together using an image labeling technique and to compute the size of each skin pixel group. Skin pixel groups with a size less than 1% of the largest skin pixel group will then be treated as noise and thus eliminated. The assumption used is that the largest skin candidate region is valid and it is used a basis to compare against other skin candidate regions. Another assumption made is that false detection happens at a very much smaller scale compared to the largest valid skin region. A median filtering process is then used to eliminate noise within the skin pixel region. This is followed by an image dilation process which helps reconstruct the face region when pixels within the region are discarded unintentionally.

a) Connected Component Labeling: Since false detection could occur for objects with similar color to the skin color, subsequent filtering process based on the size of skin candidate region is used to eliminate those unwanted objects. In order to determine the suitable size of the structuring element or filter for the median filtering and image dilation process, the size of the largest skin candidate region, $Area_{max}$ is first taken as the basis. $Area_{max}$ is determined by counting the number of pixels belonging to the largest connected component. All other skin candidate regions are then parsed to determine their respective sizes. Skin candidate region which is less than one percent of the largest skin candidate region is then rejected.

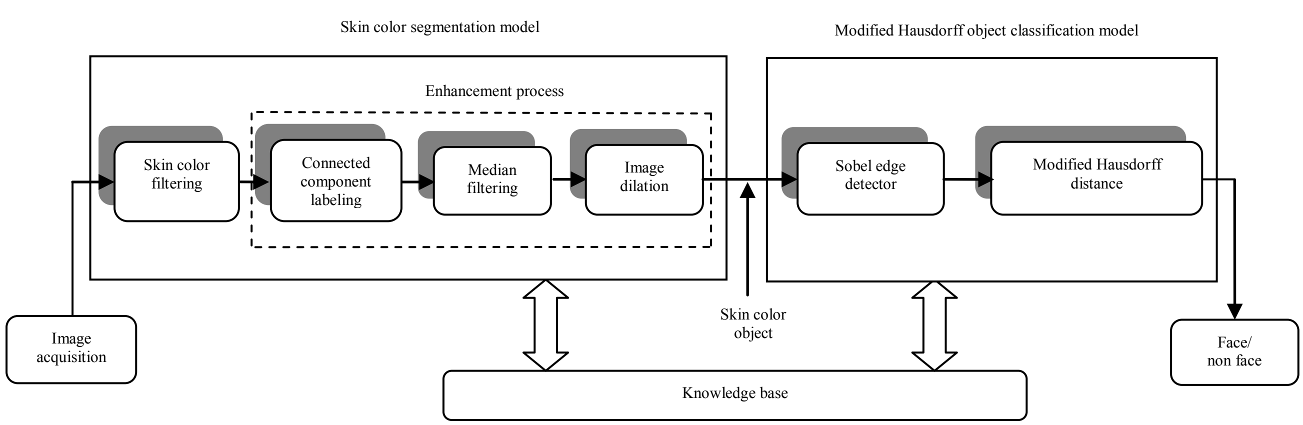


Fig. 1. The proposed face detection system.

b) Median Filtering: A median filter provides a powerful method for smoothing signals and images by repeating the median filtering the root pixel, which is invariant to further filtering is found [11]. This convergence property of repeated median filtering has been used in some signal processing applications, such as speech processing [12] and image coding [13]. A square structuring element of an equivalent area to one percent of the largest skin candidate region, $Area_{max}$ is used for the median filtering process. The size of the square structuring element is calculated as shown:

$$W_{median} = \sqrt{Area_{max}} \quad (2)$$

where W_{median} is the width of the median filter structuring element.

C) Image Dilation: Image dilation is applied to the median filtered image to reconstruct the skin color regions when certain pixels within the actual skin color regions are unintentionally discarded in the skin color filtering process. It also carries the objective of reconnecting dissociated skin color regions. The basic effect of the operator on a binary image is to gradually enlarge the boundaries of regions of foreground pixels. As a result, the areas of foreground pixels grow in size while holes within those regions become smaller. The dilation operator takes two pieces of data as inputs. The first is the image, which is to be dilated. The second is a set of coordinate points known as a structuring element.

B. *Modified Hausdorff Object Classification model*

Hausdorff distance is the maximum distance of a set to the nearest point in the other set. More formally, Hausdorff distance from set *A* to set *B* is a maximin function, defined as:

$$h(A,B) = \min\{\max_{a\in A}\{\min_{b\in B} d(a,b)\}\} \quad (3)$$

where a and b are points of sets *A* and *B* respectively, and d(a, b) is any metric between these points. For simplicity, will take d(a, b) as the Euclidian distance between *a* and *b*.

Owing to the nature of the Hausdorff distance which only takes into account the maximum distance between the two sets of points, any outlier or noise could easily lead to false detection of the face should face detection is only based on this measure. A modified version of the classical Hausdorff distance which is also known as the modified Hausdorff distance is therefore adopted. The modified Hausdorff distance measure also known as mean Hausdorff distance measure is formulated as follows:

$$h_{mean}(A, B) = \frac{\sum \max_{a\in A}\left\{\min_{b\in B} d(a,b)\right\}}{n_A} \quad (4)$$

where n_A is the total number of points in set *A*. By taking into account the mean distance between the two sets of points instead of the maximum distance only, the outlier effect could be significantly reduced.

The modified Hausdorff object classification module aims to identify the face region with higher confidence. Basically, it is divided into two stages, which are Sobel edge detection and modified Hausdorff distance matching. The Sobel edge detector will detect edges of the face candidate region along with the facial features. Modified Hausdorff distance is then used to compare the face template with the image to verify whether it is a face.

1) Sobel Edge Detector: In Hausdorff distance object matching, outline of the object is sufficient for successful comparison. Therefore, edge extraction is required to successfully extract the outline and any details in the image. Sobel edge detector fulfils this role by producing single pixel thick edges of the skin regions and outlines of the facial features. Sobel edge extraction will then produce an image, *J(x, y)* which only consists of the boundary of each region based on the formula below:

$$J(x,y) = \begin{cases} 1 & \text{when } |G| \geq \tau_{Threshold} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where $\tau_{Threshold}$ is obtained by Matlab function known as *edge* which is automatically selects the optimum threshold value and G is the Sobel convolution mask.

2) Modified Hausdorff Distance: In order to compute the modified Hausdorff distance between the template and input images, distance transform is first performed on the face template image. A matrix of the same dimension as the template image is created. An AND operation is performed between the distance transform matrix of the template image and the input image. The non-zero distances in the product matrix are then accumulated and divided by the total number of non-zero pixels of the product matrix. The result is the mean forward Hausdorff distance between the template image and input image. The same procedure is repeated for the input image where its distance transform matrix is created. The result of it is known as the reverse mean Hausdorff distance. The mean Hausdorff distance will then be the maximum between the forward and reverse mean Hausdorff distance. Thresholding is then needed to decide the adequate threshold value which acts as a cut off point between face and non face. Using a series of frontal face from the face database and non face image, the mean Hausdorff distance is decided based on the mean Hausdorff distance generated.

III. EXPERIMENTAL RESULTS

A. *Formation of Skin Color Filter*

The skin color filter is modeled as a bivariate normal distribution using two hundred images from the face database manually. Fig. 2 illustrates part of the images used for the formation of skin color filter. There is however a few criteria used in selecting the proper images used for the formation of the skin color filter. First, the object should be taken under ambient lighting where no excess lighting is used as reflection on the skin will distort its actual value in the YCrCb color space. Secondly, object chosen should have minimal obstruction on the face. Using samples of images, the mean matrix and the covariance matrix are found to be:

$$\mathbf{m}=\begin{bmatrix}\mu_{Cr}\\ \mu_{Cb}\end{bmatrix}=\begin{bmatrix}153.1373\\ 99.6061\end{bmatrix} \quad (6)$$

$$\mathbf{C}=\begin{bmatrix}144.2155 & -35.0688\\ -35.0688 & 178.7794\end{bmatrix} \quad (7)$$

Images in Fig. 2 will be used to illustrate the result of each step taken in methodology from skin color segmentation leading to the detection of face based on modified Hausdorff distance measure.

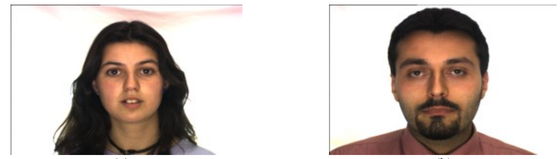
B. Skin Color Segmentation

Skin color segmentation consists of skin color filtering, thresholding, connected component labeling, median filtering and image dilation steps.

1) Skin Color Filtering: As shown in Fig. 3, pixels in white represent higher probability of that pixel belonging to a skin object while darker pixel especially pixels outside the actual skin region have lower probability of belonging to skin object. Some of the pixels which belong to the skin region still have lower probability of belonging to a skin object. Some facial features such as eyes, eye brows, and lips which have significantly different color from human skin color also have lower probability as shown by the darker pixels at the locations of these features.

2) Thresholding: Thresholding determines the suitable threshold value to be used for the cut-off value between skin color pixel and non skin color pixel. In the experiment, it found that the probability threshold value, $\tau_{\text{threshold}}$ is dependent on the input image itself and that it is not possible to assign a single threshold value which will be applicable for every input image. Therefore, an iterative process is used whereby the starting threshold value is 0.9, a step of 0.05 is then decreased from the starting threshold value until the total pixels change in the resulted image began to stabilize. Most of the face area pixels are identified as skin where they are represented by white while other pixels below the threshold value are represented as black. As shown in Fig. 4, facial features such as eyebrows, eyes and lips and even moustache which have significantly different color from human skin are filtered out during this stage. In Fig. 4 (a), (b), under individual image is the threshold value at which the result is obtained using an automatic thresholding process. This is the value when the skin filtering result begins to stabilize. An optimum threshold value is important in the sense that significant amount of skin color pixels are identified without severe false detection on the non skin color pixels.

3) Connected Component Labeling: Darker blue color is used to denote the major connected component of the face while the smaller connected component of different color can be seen. The lesser number of different connected components derived, the better is the skin color filtering process. As seen in the images in Fig. 5, skin color pixels which were originally connected are separated from each other after the skin color filtering process. In Fig. 5 (a), the face region and the neck region are separated from each other due to the fact that part of the neck region which is under the shadow of the



(a) (b)

Fig. 2. Sample images from face database.

(a) (b)

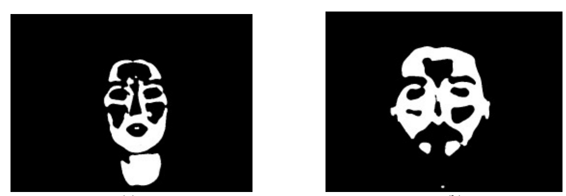
Fig. 3. Skin color distribution.

Threshold value 0.7 Threshold value 0.55

Fig. 4. Thresholding result for images.

(a) (b)

Fig. 5. Connected component labeling for images.



(a) (b)

Fig. 6. Median filtering results for images.

(a) (b)

Fig. 7. Image dilation results for images.

face region are considered as non skin color pixels under the skin color filtering process.

4) Median Filtering: In the experiment, median filter is used to smoothen the segmented skin color regions. In Fig. 6, median filtering removes some of the connected components not connected to the major connected components and creates a smoother outline for all the face regions. The smaller connected components in the forehead region are joined into the major face region after median filtering process. In Fig. 6

(b), median filtering process successfully removes most of the connected components outside the actual face region.

5) Image Dilation: Image dilation is performed on the images in Fig. 6 to reconstruct the actual skin color region when part of the skin color region is unintentionally discarded in skin color filtering and median filtering. Image dilation helps to close back some of the empty regions within the skin color region which is evident in all the images used. In some instances image dilation helps to reconnect the skin color regions which were dissociated from each other. In Fig. 7(a) in particular, the forehead region which gets separated from the rest of the face region after the median filtering operation is reconnected after image dilation operation. In Fig. 7 (a) and (b), the neck region and the face region which is still separated from each other in the median filtering process are joined back into a single region after the image dilation process.

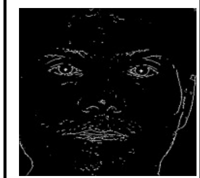
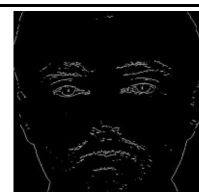
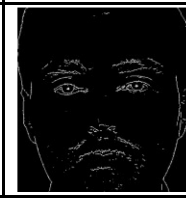
C. Modified Hausdorff Object Classification

1) Sobel Edge Detector: The outlines of the candidate skin color objects along with any possible facial features within the skin color objects can be created using the Sobel edge detector. Edges of the images which will be further evaluated by the modified Hausdorff classifier is obtained from part of the images in Fig. 3. These parts which are also known as the face candidate regions. As shown in Table I, the face candidate regions in the Fig. 2 are operated on by the Sobel operator in the *x*-direction, *y*-direction and the final column illustrates the result of an OR operation between the results of the previous two columns. It is easily noticeable that facial features are sensitive to different types of Sobel operators. Horizontal facial features such as lips and eye brows can be easily extracted using the Sobel operator in the *x*-direction by not the Sobel operator in the *y*-direction. Noses on the other head are easily detected by the Sobel operator in *y*-direction instead.

Facial feature such as eyes are picked up by both of the Sobel operator as they consist of horizontal and vertical elements. Therefore, an OR operation between the results of the two kinds of Sobel operators are able to retain most of the facial features which are of utmost importance in modified Hausdorff object matching. These features are then compared with those of the face model to calculate the modified Hausdorff distance.

2) Modified Hausdorff Object Matching: Modified Hausdorff distance is used as the basis for the object matching method. Thresholding is needed to decide the adequate threshold value which acts as a cut-off point between face and non face. Using a series of frontal face from the face database and non face image, the mean Hausdorff distance is decided based on the mean Hausdorff distance generated. In the end, it is found that the mean Hausdorff distance of 18 is an adequate threshold value. In order to identify face candidate regions as face or non-face, the image in Table I (a) is chosen as the template image. Based on the template, the entire face region within the images is successfully identified using the threshold value of 18. Identified face region will then be displayed separately so that verification can be made to see if the face detection system has been accurate.

TABLE I
EDGES DETECTED USING SOBEL EDGE DETECTOR FOR IMAGES IN FIGURE 1

	G_x	G_y	G_x OR G_y
a			
b			

All of the images have modified Hausdorff distance value within the range 12 – 18. Slight obstruction of some of the facial features such as moustache and beard do not really affect the modified Hausdorff distance. In order to verify the efficiency of our algorithm, we have employ 275 different color images containing multiple faces from different races with various sizes in our experimentation.

D. Modified Hausdorff distance object matching with test images

After the face detection system is constructed using the modified Hausdorff distance object matching method, it is necessary to verify how effective it is with other images which are not from the face database. Table II shows the results of the face detection system on two test images as examples which are not from the face database. In Table II (a), the face image is detected successful where there is only one face object. In Table II (b), there are three face objects inside the image. However, only two of the images are successfully detected which are shown in the last column.

E. Comparison of face detection techniques

Comparison is made between our proposed system and the system proposed in [14]. In their system, the algorithm for face detection is based on color information. First it finds skin color pixels in an image and then skin elements are segmented into regions. Unsuitable regions are then eliminated on the basis of geometric properties of the face. Remaining regions represent faces. As illumination conditions affect the colors and this algorithm basis merely on colors, the results of face detection are influenced by illumination.

The testing set is composed of 160 images taken under four different types of illumination conditions. Four different types of color compensation algorithms are used which are: (i) Grey World (GW) algorithm; (ii) Modified Grey World algorithm (MGW) and (iii) White-patch Retinex (RET) algorithm as shown in Table III. Based on different kind of compensation algorithms used, the detection results obtained is as shown in Table III. One subset of images (40 images) was taken under standard daylight has been chosen to be the basis of comparison between the method proposed and their method.

As seen, the method proposed in this paper is superior in terms of detection rate as well false detection rate in all cases. The face detection system is able to detect faces under complex background and of different sizes and locations. Multiple faces are also able to be detected using this system.

IV. CONCLUSIONS

This paper presents a novel method for face detection system based on the modified Hausdorff distance technique using color images. The skin color segmentation process is based on a human skin color filter constructed from 200 human face images from diverse races. The face detection system is then implemented on test images with either a single face or multiple faces to measure the detection rate and the false detection rate of human faces. The final system is a face detection system which is capable of detecting human faces of different scales and from different races with a success rate of 87.5%.

TABLE II
FACE DETECTION RESULT BASED ON MEAN HAUSDORFF DISTANCE METHOD

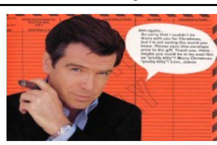
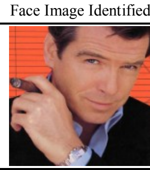
	Test Image	Face Image Identified
a		
b		 

TABLE III
FACE DETECTION COMPARSION

	Algorithm proposed in [14]				Proposed algorithm
Method	None	GW	MGW	RET	
All Faces	109				109
Detected	68	66	60	68	95
False detection	7	5	5	8	5
Detection rate	62.39	60.55	55.05	62.39	87.5
False detection rate	6.42	4.59	4.59	7.34	4.59

ACKNOWLEDGMENT

This Project is supported by the General Peoples Committee for Higher Education, Libyan Arab Jamahiriya and the International Science Linkages established under the Australian Government's innovation statement, Backing Australia's Ability.

REFERENCES

[1] C. Liu, "A bayesian discriminating features method for face detection," *IEEE Trans. Pattern Analysis and Machine Intelligence.,* vol. 25, no. 6, pp. 725-740, June 2003.

[2] H. Kim, S. Lee, and N. Cho, "Rotation-invariant face detection using angular projections," *Electron. Lett.* , vol. 40, no. 12, pp. 726- 727, June 2004.

[3] Y. Suzuki, and T. Shibata, "Multiple-clue face detection algorithm using edge-based feature vectors," in *Proc. ICASSP*, May 2004, pp. 737-740.

[5] Z. Zhijia, H. Shabai, and S. Zelin, "A fast strategy for image matching using Hausdorff distance," in *Proc. IEEE International Conference on Robotics, Intelligent Systems and Signal Processing,* Changsha, China, Oct. 2003, vol. 2, pp. 915-919.

[4] P. Lekshmi, S. Kumar, and D. Vidyadharan, "Face detection and localization of facial features in still and video images," in *Proc. First International Conference on Emerging Trends in Engineering and Technology (ICETET'08),* July 2008, pp. 95-99.

[6] Y. Labit, and J. Mazel, "HIDDEN: Hausdorff distance based intrusion detection approach dedicated to networks," in *Proc. The Third International Conference on Internet Monitoring and Protection (ICIMP' 08),* June 2008, pp. 11-16.

[7] Y. Gao, "Efficiently comparing face images using a modified Hausdorff distance," *IEE Proceedings online. Vision, Image and Signal Processing,* vol. 150, no. 6, pp. 346- 350, Dec. 2003.

[8] P. Kumar, K, Sengupta, A. Lee, "A comparative study of different color spaces for foreground and shadow detection for traffic monitoring system," in Proc. *The IEEE 5th International Conference on Intelligent Transportation Systems,* Sep 2002, pp. 100-105.

[9] M. Asmare, V. Asirvadam, L. Iznita, "Color Space Selection for Color Image Enhancement Applications," in *Proc. International Conference on Signal Acquisition and Processing (ICASP'09),* April 2009, pp. 208-212.

[10] A. Martinez and R. Benavente, "The AR face database" http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html

[11] A. Burian, P. Kuosmanen, "Tuning the smoothness of the recursive median filter," *IEEE Trans. Signal Processing.,* vol. 50,no. 7, pp.1631-1639, Jul 2002.

[12] L. R. Rabiner,M. R. Sambur, and C. E. Schmidt, "Application of a nonlinear smoothing algorithm to speech processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 552–557, Dec. 1975.

[13] G. R. Arce and N. L. Gallagher Jr., "BCT image coding using median filter roots," *IEEE Trans. Commun.*, vol. COMM-31, pp. 784–793, June 1983.

[14] J. Kovac, P. Peer, F. Solina, "Illumination independent color-based face detection," in *Proc. The IEEE 3rd International Symposium on Image and Signal Processing and Analysis (ISPA'03),* Sep 2003, Vol. 1, pp. 510- 515.