International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:2, No:5, 2008

Adaptive Gaussian Mixture Model for Skin Color Segmentation

Reza Hassanpour, Asadollah Shahbahrami, and Stephan Wong

***Abstract*—Skin color based tracking techniques often assume a static skin color model obtained either from an offline set of library images or the first few frames of a video stream. These models can show a weak performance in presence of changing lighting or imaging conditions. We propose an adaptive skin color model based on the Gaussian mixture model to handle the changing conditions. Initial estimation of the number and weights of skin color clusters are obtained using a modified form of the general Expectation maximization algorithm, The model adapts to changes in imaging conditions and refines the model parameters dynamically using spatial and temporal constraints. Experimental results show that the method can be used in effectively tracking of hand and face regions.

***Keywords*—Face detection, Segmentation, Tracking, Gaussian Mixture Model, Adaptation.

I. Introduction

HUMAN body, face or hand movements are very complicated in comparison to artificial articulated objects. This means that all available cues which can narrow the search space in tracking or detection systems should be considered. Among these features, spatial, temporal and textural features are the most important ones. The simplicity and obviousness of skin color as an effective cue for hand and/or face segmentation and tracking has caused many researchers to develop methods based on this feature [15], [1], [2], [18], [14]. The most important parameter in developing skin color based segmentation methods lies in choosing the color space and the model used for representing the distribution of the skin color values. A later step involving processing the segmentation results can be affected drastically from the color distribution model parameters and the space used, so an efficient segmentation is the key feature in the successful implementation of detection and tracking systems. Pixel-based skin color segmentation on the other hand is very sensitive to the environmental effect such as noise and illumination. Parametric techniques which use statistical models have been more or less successful in reducing some of these impacts but since the stochastic models are based on static parameter computation, temporal effects such as gradual change of illumination are either not dealt with or had a negative impact like decreased accuracy of the model. The method proposed here is based on the Gaussian Mixture Model (GMM) with the exception that the training step is a dynamic one which makes it possible for the model to adapt itself with the environmental changes. The spatial and temporal constraints are also applied to the model adaptation which makes the method applicable to motion detection in video stream compression as well. These considerations put our algorithm in both pixel-based and region based groups of methods.

The remainder of this paper is organized as follows. Section II describes the different color space conversions that are used in the skin color modeling. Section III discusses the Gaussian mixture model followed by Section IV describes the modified Expectation maximization algorithm. Section V explains model adaptation. Experimental results are discussed in Section VI. Finally, conclusions are given in Section VII.

All authors are with the Computer Engineering Laboratory, Delft University of Technology, Delft, The Netherlands. Email: reza@dutepp0.et.tudelft.nl, A.Shahbahrami@TUDelft.nl, J.S.S.M.Wong@ewi.tudelft.nl. Reza Hassanpour is also with Computer Engineering Department, Çankaya University, Ankara Turkey, reza@cankaya.edu.tr. Asadollah Shahbahrami is also with Department of Computer Engineering, Faculty of Engineering, University of Guilan, Rasht, Iran.

II. Skin Color Modeling

Pixel-based skin color detection methods aim at introducing a tool for measuring the distance of each pixel color to skin color tones. The color itself can be represented in many different ways among which the most widely used ones are summarized below.

A. RGB and Normalized RGB

The RGB color space has been widely used for processing and storing digital image data. This model describes each color as a weighted combination of three base components Red, Green and Blue. However, high correlation between components and mixing luminance with chromaticity makes it very sensitive to changes in imaging conditions such as lighting. Normalized RGB tries to reduce the dependence of each component to the brightness of the pixel by normalizing each component using:

$$r = \frac{R}{R+G+B} \qquad g = \frac{G}{R+G+B} \qquad b = \frac{B}{R+G+B}. \tag{1}$$

In fact, since the sum of the normalized components is equal to 1, the third component does not hold any significant information. The simplicity of these color spaces has been the main reason for their popularity in skin color detection [8], [4].

B. Hue Saturation Lightness Model

Hue Saturation Lightness (HSL) model describes color with dominant color (Hue), colorfulness in proportion to the brightness (Saturation), and the amount of luminance (Lightness). The most important characteristic of the model is its explicit discrimination of luminance from chrominance. This makes the model insensitive to brightness at white color and ambient light; the properties that have attracted many researchers to put their skin color detection works on this model [19], [11], [16], [3]. H, S, and L components are computed using the following equations.

$$H = \arccos \frac{\frac{1}{2}((R-G)+(R-B)}{\sqrt{(R-G)^2+(R-B)(G-B)}}. \tag{2}$$

$$S = 1 - 3\frac{\min(R,G,B)}{R+G+B}. \tag{3}$$

$$V = \frac{1}{3}(R+G+B). \tag{4}$$

The model, however, has the disadvantage of being discontinuous at points where the brightness is very low (very dark points).

C. YCbCr

The YCbCr color space was developed as part of ITR-R BT.601 during the development of a world wide digital component video standard which is commonly used by European television studios. YCbCr separates luminance from chrominances in RGB values using a linear transform consisting of a weighted some of the three components. The simplicity of the transform which is given in Equation (5) and the explicit separation of luminance have made the model very attractive for skin color detection [5], [10], [17].

$$Y = 0.299R - 0.587G - 0.114B. \tag{5}$$

$$C_b = R - Y. \tag{6}$$

$$C_r = B - Y. \tag{7}$$

D. Non-parametric Color Distribution Modeling

This group of methods estimates the probability of a color value from the training data without defining any explicit model. The probability of a color value being a skin color is estimated by quantizing the histogram of the data. A conditional probability based on Bayes classifier is commonly used for separating skin and non-skin pixels [8], [19], [5], [10]. The Bayes classifier rule is given as:

$$P(skin|c) = \frac{P(c|skin)P(skin)}{P(c|skin)P(skin) + P(c|-skin)P(-skin)}. \tag{8}$$

where $P(skin|c)$ is the probability of being a skin pixel given the color value c.

E. Parametric Color Distribution Modeling

Parametric color distribution tries to describe the chrominance feature space using a statistical model. Obviously the key problem here is finding the best model and estimating its parameters. The estimation should reasonably well fit the training data where the goodness of fit depends on the shape of the distribution and therefore the color space used [17], [9].

F. Single Gaussian Model

A multivariate normal distribution of a D-dimensional random variable x is defined as:

$$N(x;\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma|^{\frac{1}{2}}} \exp[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)]. \tag{9}$$

where μ is the mean vector and Σ the covariance matrix of the normally distributed random variable x. The model parameters are estimated from the training data using the following equations.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} c_i. \tag{10}$$

$$\Sigma = \frac{1}{n-1}\sum_{i=1}^{n}(c_i-\mu)(c_i-\mu)^T. \tag{11}$$

Either the $p(c|skin)$ probability or the Mahalanobis distance from the c color vector to mean vector μ, given the covariance matrix Σ can be used to measure the similarity of the pixel with the skin color [7].

III. Gaussian Mixture Model

Despite the fact that the single Gaussian models have been successfully used to represent features and discriminate between different classes in many practical problems, the assumption of single component requires a single basic class which smoothly varies around the class mean. This requirement assumes a unimodal distribution which may cause intolerable error in estimation and discrimination. A better approximation can be obtained when the values are generated by one of the several randomly occurring independent sources [13], [12]. In this case the distribution function is a multimodal one which can be estimated using a finite number of mixed Gaussian or a Gaussian mixture model. The GMM probability density function can be defined as a weighted sum of Gaussian as [6]:

$$P(x;\theta) = \sum_{i=1}^{N} \alpha_i G(x;\mu_i,\alpha_i). \tag{12}$$

where α_i is the weight of i^{th} component. The weight can be interpreted as *a priori* probability that a value of the random variable belongs to the i^{th} group. G is a Gaussian probability density function with parameters μ and σ. In addition, x is a sample input and N is the number of components. The parameter list of the Gaussian mixture model probability density function is given by:

$$\theta\{\alpha_i,\mu_i,\sigma_i\} \quad for \quad i = 1..N. \tag{13}$$

Model parameter estimation is performed using a well known iterative method called Expectation Maximization (EM) which assumes that the number of components is known beforehand. Here we have introduced a modified form of this algorithm which can change the initial estimation of the number of components and is more suitable for our application.

IV. Modified Expectation Maximization Algorithm

Considering that skin colors samples from different ethnic groups generates a multimodal random variable, a finite mixture model is used to approximate the pdf. Here we have assumed that a Gaussian form is sufficient for each single source. Suppose $\boldsymbol{X}$ is the set of independent samples drawn from a single distribution by $P(x;\theta)$ where θ is the list of parameters of the probability distribution function (pdf). Maximum Likelihood function given in Equation (12) estimates the set of parameters which describes the sample data best.

$$L(X;\theta) = \prod_{n=1}^{N} P(x_n;\theta). \tag{14}$$

The EM algorithm is used for calculating the distribution parameters using maximum likelihood. The algorithm can also be used to handle cases where an analytical approach for maximum likelihood estimation is infeasible, such as Gaussian mixtures with unknown and unrestricted covariance matrices and means. The EM algorithm includes two steps:

1) Expectation or E step
   This step consists of forming the function given in the following equation.

$$Q(\theta;\theta^i) = E[\ln L(X;\theta)|X;\theta^i]. \tag{15}$$

   where θ^i is the current list of the parameters and θ is a variable which will return the new parameters. Equation (15) defines the relation using the logarithm of the likelihood function. Because of the monotonicity of the logarithm function, instead of the likelihood function sometimes its logarithm called log-likelihood is used which is simpler to deal with.
2) Maximization or M step
   This step involves maximizing $Q(\theta;\theta^i)$ with respect to θ as shown in the following equation.

$$\theta^{i+1} = \arg\max_{\theta} Q(\theta;\theta^i). \tag{16}$$

The modified EM algorithm starts with an initial number of component and an initial parameter list. The algorithm consists of two stages. The first stage follows the general EM algorithm with a fixed number of components. Stage two tries all elements of the components. If the distance of an element to the mean of its component is greater than $2.5\sigma^2$ then a new component is created having that as the seed element. A post-processing step removes components with number of elements less that a threshold. The members of these components are considered as outliers and ignored.

V. Model Adaptation

The parameters of the GMM model is initially estimated using a set of training points. However, the imaging conditions such as lighting may change as time passes. The model parameters are regularly adapted to cope with these changes. The adaptation is based on two important heuristics:

1) The pixels in a segmented connected component in a frame should maximize the same component in the mixture of *pdf*s. This fact imposes a restriction based on the spatial proximity of the skin pixels.
2) The pixels from a segmented connected component should maximize the same component in the mixture of *pdf*s in the consecutive frames. This rule imposes temporal restriction on classification.

Violations from the above mentioned heuristics are due to changing conditions which should be compensated for by adapting model parameters. The proposed algorithm is as follows:

- Segmentation
  The segmentation step aims at separating and classifying all skin color pixels in connected components. A Gaussian Component Map (GCM) is also defined and initialized for all frame pixels. An entry in this map shows to which Gaussian component a skin color belongs. If a pixel is not classified as a skin color then the corresponding place in GCM is initialized to -1 as shown in Algorithm 1.

Algorithm 1 Segmentation Algorithm

for each pixel x **do**
  if P(x,θ)>Threshold **then**
    Classify the pixel as skin color
    c=Maxarg(G(x,μ_i,σ_i)), i=1..N.
    Assign c to the Gaussian Component to which the current pixel belongs in the GCM.
  else
    Assign -1 as the corresponding Gaussian component in the GCM.
  end if
end for
Find all connected components in the skin colored pixels.

- Model adaptation using spatial restriction.
  This step assumes that the skin colored pixels in a connected component should belong to the same Gaussian component. However, due to changing illumination conditions this restriction may not be satisfied and therefore the model parameters should experience a gradual drift to incorporate the new conditions. Algorithm 2 shows the procedure.
- Model adaptation using temporal restriction.
  We assume that a segmented component in a frame will experience a slight displacement in the following frame but no major changes should we have in the Gaussian component map obtained in step a and updated in the second step. This is given in Algorithm 3.

Updating model is performed as follows:

$$\mu_c(i) = [1-\alpha]\mu_c(i-1) + \alpha I. \tag{17}$$

$$\sigma_c^2(i) = [1-\alpha]\sigma_c^2(i-1) + \alpha[\mu_c(i)I]^2. \tag{18}$$

where $\mu_c(i)$ and $\sigma_c^2(i)$ are the mean and standard deviation of the component c after being updated. In addition, I is the pixel value which triggers the adaptation process. The number

Algorithm 2 Segmentation Algorithm

for i in the segmented components **do**
  for j in the skin colored pixels of i **do**
    Define a window W with height and width equal to H centered at pixel j.
    D ← DominantSkinColoredComponent(W)
    if j does not belong to the Gaussian component D **then**
      Update model by adapting component D
    end if
    if Gaussian component of j does not match with its corresponding value in GCM **then**
      Update GCM
    end if
  end for
end for

Algorithm 3 Segmentation Algorithm

for all connected components in frame i **do**
  Track the component in frame i+1 using Mean Shift operator.
  for all pixels j in the current connected component **do**
    Find d=j is Gaussian component from FCM
    Compute c= j is Gaussian component from equation 10
    if c <> d **then**
      Update model by adapting the current component.
    end if
  end for
end for

of components and their weights are not updated because firstly these parameters have been obtained from a large set of training samples and secondly we assume the adaptations are due to changing imaging conditions such as illumination while the initial classification is based on the ethnic characteristics of the skin color values.

VI. Experimental Results

The initial training of the Gaussian components is carried out using a data set of skin color pixels that includes 127,352,563 pixels. The training images have cautiously been selected to cover almost all ethnic groups and imaging conditions. The training images were manually segmented before parameter estimation. Figure 1 shows a subset of the training images and Figure 2 shows them after being manually segmented. The color space used is YCbCr. To reduce the effect of illumination, Y channel has not been considered through out the segmentation process. We start the initial EM stage with five components. This assumption is based on our intuitive classification of the skin samples using their appearance color. The second stage of EM algorithm tries to optimize the number of components. According to the results of this stage, the optimum number of components is seven. The adaptive algorithm uses the results of this stage for segmentation.
Figure 3 shows a sequence of frames indicating changes in lighting condition. Some pixels from the first frame have been

Fig. 1. Samples from training data.

Fig. 2. Manually segmented training data.

compared to their correspondences in the last frame in the sequence given in frame 3. Four small regions have been

Fig. 3. A video sequence with changing lighting condition.

marked with red circles in the video sequence to show the effect of changing illumination. The YCbCr values and the changes in each component of these pixels are given in Table I. Despite having the maximum changes in Y component, illumination changes may cause variations in Cb and Cr components which can result in misclassification. We tested and compared the proposed algorithm with the performance of skin color segmentation without parameter adaptation. Our testing stage utilizes three sets of testing image sequences obtained from 7 video sequences. The first set includes the regions with skin colors which have been manually segmented. The second set is the complement of the first set where the images include no

International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:2, No:5, 2008

TABLE I
ILLUMINATION EFFECT ON THE PIXEL VALUES AND THEIR CLASSIFICATIONS.

	Image 1 (Y,Cb,Cr)	Image 2 (Y,Cb,Cr)	Image 3 (Y,Cb,Cr)
Forehead	(114.4, -7.5, 18.1)	(147.2, -4.8, 13.8)	(32.8, 2.7, -4.3)
Cheek	(122.7, -10.8, 18.2)	(146.8, -7.3, 22.8)	(24.1, 3.5, 4.6)

skin pixel. Finally, the third set is the set of the original video frames. The first and the second sets are used as the ground truth for the results we obtain by applying the tests to the third set. The testing procedure starts after initial training of the model and counts the number of false positive and negative responses generated by both adaptive and non-adaptive GMM. Figure 4 compares the results of segmentation using adaptive and non-adaptive GMM model. Our experimental results show that the proposed method has a better performance compared to the non-adaptive GMM both in reducing the number of false-positive and false-negative cases. Figure 5 shows the comparative performance of the two methods.

VII. CONCLUSION

An adaptive skin color segmentation algorithm based on Gaussian Mixture Model (GMM) has been proposed which can adapt the model parameters to cope with the changing imaging conditions such as lighting and noise. The algorithm considers the higher level *a-priori* knowledge of spatial and temporal relationship between the pixels in a video sequence. The experimental results show the superiority of the proposed algorithm compared to non-adaptive parametric models. Further improvement can be achieved by incorporating application specific information such as human motion models or face-hand detectors. The proposed method may also be used in surveillance systems and motion detectors.

Fig. 4. Skin color regions segmented using non-adaptive GMM (left), proposed method (middle), and the original image (right).

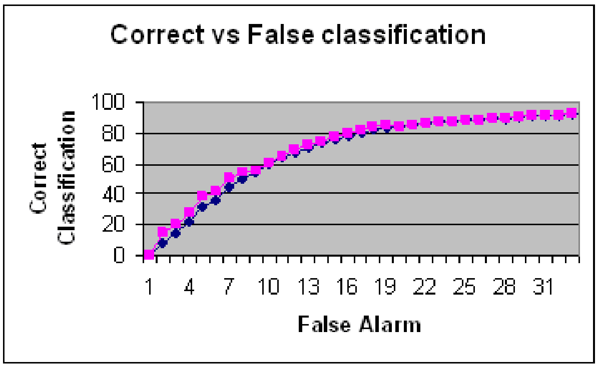

Fig. 5. Performance comparison of the non-adaptive and proposed method with respect to the ratio of correct classification to false alarms. The brighter curve shows the proposed method performance.

Reza Hassanpour received his BSc, MSc and PhD degrees in computer engineering from Shiraz University, Shiraz-Iran, Tehran Polytechnic University, Tehran-Iran and Middle East Technical University, Ankara-Turkey in 1995,1998 and 2003 respectively. In 1999 he joined the Department of Computer Engineering at Cankaya University, Ankara-Turkey as a faculty member. He has recently started working on hand gesture recognition algorithms as a PostDoc researcher in the Computer Engineering Department of Delft University of Technology. His main fields of interest are 3D machine vision, pattern recognition and computer graphics.

Asadollah Shahbahrami received the M.Sc degree in Computer Engineering-Machine Intelligence, from Shiraz University, Shiraz, Iran, in 1996. He has worked as a member of faculty staff in the Electrical Engineering Department at the University of Guilan, Guilan, Iran, for 7 years from 1996 to 2003. In January 2004, he joined the Faculty of Electrical Engineering, Mathematics, and Computer Science (EEMCS) at Delft University of Technology, The Netherlands, as a full time Ph.D student under advisor of Prof. Stamatis Vassiliadis and Dr. Ben Juurlink. His research interests include computer architecture, image and video processing, multimedia information retrieval, multimedia instructions set design, and SIMD programming. He is a member of the HiPEAC, IEEE, and ACM.

Stephan Wong received his PhD degree in 2002 from the Electrical Engineering Department at Delft University of Technology (TU Delft), The Netherlands. He is currently working as an assistant professor in the Computer Engineering Laboratory at TU Delft. He has considerable experience in the design of embedded media processors. He has also worked on microcoded FPGA complex instruction engines and the modeling of parallel processor communication networks. His research interests include embedded systems, multimedia processors, complex instruction set architectures, reconfigurable and parallel/distributed processing, microcoded machines, and network processors. He is a member of the IEEE, HIPEAC, and ACM.

REFERENCES

[1] A. A. Argyros and M. I. A. Lourakis. Real-Time Tracking of Multiple Skin-Colored Objects with a Possibly Moving Camera. In *Proc. European Conference on Computer Vision (ECCV)*, pages 368–379, 2004.

[2] A. A. Argyros and M. I. A. Lourakis. Three-Dimensional Tracking of Multiple Skin-Colored Regions by Moving Stereoscopic System. *Applied Optics*, 43(2), January 2004.

[3] S. Birchfield. Elliptical Head Tracking Using Intensity Gradients and Color Histograms. In *Proc. on Computer Vision and Pattern Recognition*, pages 232–237, 1998.

[4] J. Brand and J. Mason. A Comparative Assessment of Three Approaches to Pixel Level Human Skin Detection. In *Proc. IEEE Int. Conf. on on Pattern Recognition*, pages 1056–1059, 2000.

[5] D. Chai and A. Bouzerdoum. A Bayesian Approach to Skin Color Classification in YCbCr Color Space. In *Proc. 10th IEEE Conf. on Region*, pages 421–424, 2000.

[6] B. S. Everitt and D. J. Hand. *Finite Mixture Distribution, Monographs on Applied Probability and Statistics*. Chapman and Hall, 1981.

[7] R. L. Hsu, M. Abdel-Mottaleb, and A. K. Jain. Face detection in color images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 24(5):696–706, May 2002.

[8] M. Jones and J. M. Rehg. Statistical Color Models with Application to Skin Detection. In *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 274–280, 1999.

[9] J. Lee and S. Yoo. An Elliptical Boundary Model for Skin Color Detection. In *Proc. Int. Conf. on Imaging Science, System and Technology*, 2002.

[10] C. Liu. A bayesian discriminating features method for face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(6):725–740, June 2003.

[11] S. J. McKenna, S. Gong, and Y. Raja. Modelling Facial Colour and Identity with Gaussian Mixtures. In *Proc. on Pattern Recognition*, pages 1883–1892, 1998.

[12] K. Nigam, A. McCallum, S. Thrun, and T. Mitchell. Text Classification from Labeled and Unlabeled Documents Using EM. *Machine Learning Special Issue on Information Retrieval*, 39(2):103–134, May-June 2000.

[13] N. Oliver, A. Pentland, and F. Berard. Lafter: Lips and Face Real Time Tracker. In *Proc. Computer Vision and Pattern Recognition*, pages 123–129, 1997.

[14] P. Perez, C. Hue, J. Vermaak, and M. Cangnet. Color-Based Probabilistic Tracking. In *Proc. European Conference on Computer Vision (ECCV)*, pages 661–675, 2002.

[15] M. Sadeghi, J.V. Kittler, A. Kostin, and K. Messer. A Comparative Study of Automatic Face Verification Algorithms on the BANCA Database. In *Proc. Int. Conf. on Audio and Video Based Person Anthentication*, pages 35–43, 2003.

[16] L. Sigal, S. Sclaroff, and V. Athitsos. Estimation and Prediction of Evolving Color Distributions for Skin Segmentation Under Varying Illumination. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1883–1892, 2000.

[17] J. Terrillion, M. Shirazi, H. Fukamachi, and S. Akamatsu. Comparative Performance of Different Skin Chrominance Models and Chrominance Spaces for the Automatic Detection of Human Faces in Color Images. In *Proc. Int. Conf. on Face and Gesture Recognition*, pages 54–61, 2000.

[18] H. Yang, D. Kriegman, and N. Ahuja. Detecting Faces in Images: A Survey. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 24(1):34–58, 2002.

[19] D. Zarit, B. J. Super, and F. Queck. Comparison of Five Color Models in Skin Pixel Classification. In *Proc. Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems*, pages 58–63, 1999.