Face Tracking using an Adaptive Skin Color Model

L.M. Bergasa, A. Gardel, M. Mazo, M.A. Sotelo.
Departamento de Electrónica. Escuela Politécnica. Universidad de Alcalá
Campus Universitario s/n. 28805 Alcalá de Henares. MADRID. Spain
Tf:+34 91 885 6569-40 Fax: +34 91 885 6591
E-mail:bergasa@depeca.alcala.es http://www.depeca.alcala.es

Abstract

This paper presents a face tracking system that generates some commands in order to guide an electrical wheelchair for handicapped people. It works on complex color images with random background. It is not necessary neither supervised learning nor initial parameters to be introduced. It is able to find faces of different human races at any position, it is adaptive and therefore robust to light and background changes. The system has three phases: skin color segmentation, adaptive skin color tracking and face detection tracking. It has been tested with several sequences of images and the results are given. It works in real time without supervision and it also has potential in color objects segmentation problems.

Keywords: Face tracking, skin color segmentation, Gaussian functions, vector quantization learning (VQ), Kalman filter.

1. Introduction

A kind of human-machine interaction which is very interesting is the tracking of the direction where a person is looking at. This information can be required for several applications: automatic focus [Canon, 95], teleconferencing with improved visual sensation [De Silva et al.,95], faces identification in security systems [Sun&Poggio, 98], gaze driven panorama image viewer for virtual reality systems [Stiefelhagen et al. 97], lips readers [Meier et al. 97], assistance to the mobility of disable people [Heinzmann&Zelinsky 97], etc.

The Electronics Department of The University of Alcala has been working for more than 6 years on artificial means to assist the mobility of handicapped people. Nowadays, an electronic system is being developed, within SIAMO [Mazo et al.,98] project (Integral System for Assisted Mobility), in order to guide a multi-functional wheelchair for disabled or elderly people (Figure 1). This project includes an alternative guidance, by gaze direction, for cases of severe disability.

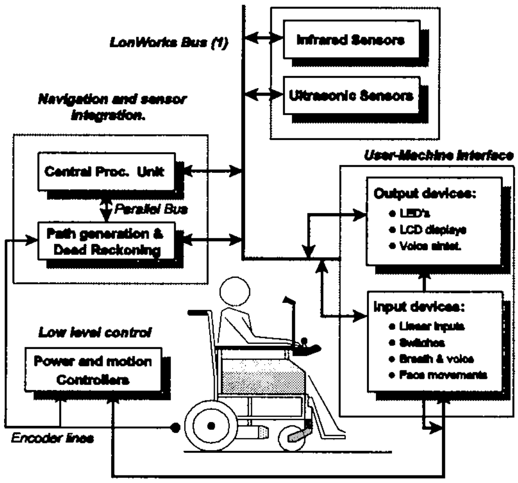

Figure 1. SIAMO Project

For doing this the system is composed of a color micro camera in front of the user, which is going to get the faces images. The images are digitized by a frame grabber and loaded in the memory of a little board based on Pentium.The gaze direction is calculated by techniques of image processing and depending on it some commands to the low level system (motor control) will be applied to make wheelchair moves properly. The system is non-intrusive and it allows the user visibility and freedom of head movements.

A person's gaze direction depends on two factors: orientation of head and eyes. While the orientation of head determines the overall direction of gaze, orientation of eyes is determining the exact gaze direction and it is limited by head position.

The tracking system which is going to be developed is a head tracking system. Next, we give a brief abstract of the different phases. In the first one the main chromaticities of an image are found using clustering techniques in a 2D normalized color space and then the skin color cluster is located. In the second phase the skin cluster is modeled by a Gaussian function. The parameters of the model are adapted by a linear combination of the known ones using the maximum likelihood criterion. In the last phase, the center position and the vertical and horizontal size of the face are estimated by a recursive estimator (Kalman filter) and the state vector is introduced in a fuzzy classification that codifies different control commands. Sections 3,4 and 5 review the phases of the system. Finally in section 6 some experimental results are provided.

2. Previous works

Previous works have applied vision-based techniques for gaze tracking like, for instance, deformable templates in b/w images [Yuille et al. 92]. Some others, as [De Silva et al.,95], have added an initialization template system based on edge pixel counting and integral projection, using two deformable circular templates. [Baluja&Pomerleau 93], from Carnegie Mellon University, have employed a neural network called ALVINN, which was designed to drive a vehicle in a road. A family of steerable Gabor Filters for detecting features points from the image, has been studied by Cipolla in the University of Cambridge[Yow&Cipolla 96]. These features points are grouped into face candidates using geometric and gray level constraints. A probabilistic framework is also employed to reinforce probabilities and to evaluate the likelihood of the candidate of a face.

On the other hand, [Yang&Waibel 96][Stiefelhagen et al. 97] have been researching into a stochastic model to characterize skin color distribution of human faces with a fixed initialization of this model parameters. However [Heinzmann&Zelinsky 97] have worked on template matching in b/w and color with a specific hardware called M.E.P. of Fujitsu. At the same time [Crowley et al. 97] was using three visual processes: blink detection, normalized color histogram matching and cross correlation. A confidence factor controls these visual processes and, depending on this factor, results are fusioned in a recursive estimator.

3. Skin segmentation using clustering

In this phase the main chromaticities of an image are found employing a clustering process, and the skin chromaticity is located. A study of different color spaces has been done in order to find the optimum for this application (RGB, normalized RGB, HSI, SCT, YQQ [Littmann&Ritter 97]) and the normalized rg space has been chosen:

$$r=\frac{R}{R+G+B} \qquad g=\frac{G}{R+G+B} \qquad (1)$$

The clustering algorithm divides the chromaticities of an image in a number of classes (k) between one and a maximum value introduced by the user (K). At each step, the k cluster centers are estimated using an approximate color histogram. After that, these centers are adjusted employing a competitive learning strategy (VQ) in a closest center sense. Finally a clustering quality factor is calculated for each topology. The process is repeated adding a new cluster center in each step until the maximum number of classes.The maximum quality factor gives the number of classes that best fits the histogram distribution. With these number of classes the skin cluster is located depending on the distance between the center of the clusters and a master skin color position. Then the skin class is modeled by a Gaussian function computing its parameters.

3.1.-Approximate color histogram

The main chromaticities in a MxM image are found using an approximated color histogram of the image. The user can choose the resolution of the histogram by specifying the number of bins, N, along each axis. Each color axis is divided up into N intervals of equal size, S, equal to the dynamic range of the color axis, P, divided by the number of bins, N. The histogram will be a NxN matrix and every bin will be initialized to zero. For every pixel, $\mathbf{x}=(x_r,x_g)$ of the image:

$$H(f_{bin}(x)) = H(f_{bin}(x)) + 1 \qquad (2)$$

$$f_{bin}(x) = (truc(\frac{x_r}{S}), truc(\frac{x_g}{S})) \qquad (3)$$

The structure of the bins in a color histogram implies an equality between colors. If the histogram has a bin size, S = P/N, then colors that are less separate than that, may fall into the same color bin and will therefore be considered as the same color. In this application only the main chromaticities in the image are sought and a 50x50 histogram has proven optimal for that task.

This system locates the initial cluster center of the k classes to be evaluated in the k biggest bins in the approximate histogram. Of course these positions aren't the ones but they are a good approximation and have a maximum error of P/2N. After that the exact centers are calculated and finally the quality factor for the k classes is evaluated.

3.2. Competitive Learning

In order to adjust the centers of the clusters a competitive algorithm proposed by Kohonen called vector quantization, VQ [Kohonen, 97] is used. A large amount of information is approximated by a vector of parameters. In this case the information is the rg color values of the pixels (**x**) and the vector contains the position of the centers of the classes. To apply VQ the number of clusters and an approximate starting center position for each must be known. In this case the algorithm starts with a good estimation of the final position. This greatly increases the ability of the algorithm to converge to the true values.

Figure 2 (a) shows a color image, and (b) shows the initial positions of the cluster centers (dark circles) and the final positions after the competitive learning (dark crosses) for the histogram of figure 2(a) with 3 classes.

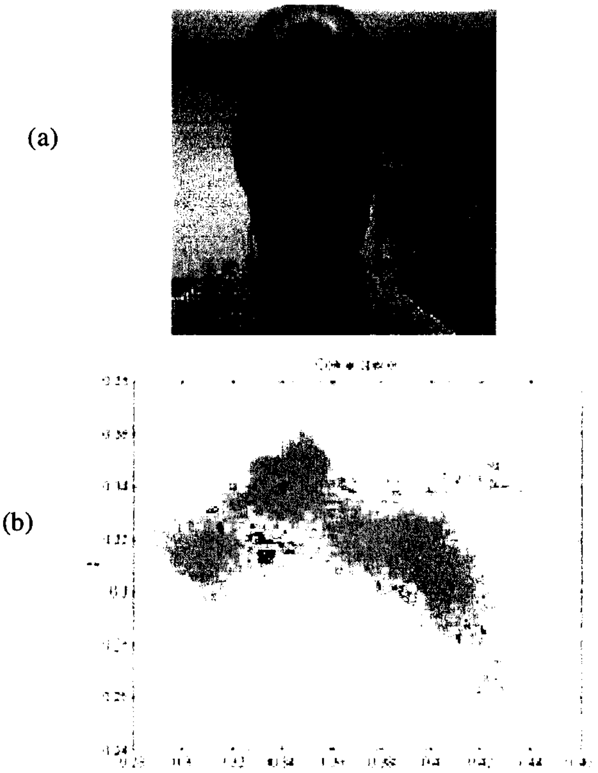

Figure 2. (a) Original Color image (b) Center of the clusters

3.3. Clustering quality factor

The clustering algorithm calculates the best approximation of the color distribution with k vectors. A quality factor is used in order to evaluate the best adjust between number of classes and color distribution. When the quality measurement reaches its maximum, the optimum number of clusters is determined. The clustering quality factor for k classes is given by (4), where tr[·] is the trace of a matrix and S_w and S_B denotes, respectively, the within-cluster and between-cluster scatter matrices.

$$F_k=tr[S_W^{-1}S_B] \qquad 1\leq k\leq K \tag{4}$$

$$S_B=\frac{1}{kM}\sum_{l=1}^{k} M_l[m_l-m_0]^T[m_l-m_0] \tag{5}$$

$$S_W=\frac{1}{k}\sum_{l=1}^{k}\frac{1}{M_l}\sum_{i=1}^{M_l}[x_i-m_l]^T[x_i-m_l] \tag{6}$$

$$m_0=\frac{1}{M}\sum_{i=1}^{M}x_i \;\; ; \;\; m_l=\frac{1}{M_l}\sum_{i=1}^{M_l}x_i \tag{7}$$

In the above equations, k is the number of clusters, M_l is the number of pixels in the l^{th} cluster, x_i is a color pixel in the l^{th} cluster, m_l is the mean of the l^{th} cluster, m_0 is the mean of all of the feature vectors and M denotes the total number of pixels to be clustered.

Figure 3 shows the clustering quality factor for the image 2(a) with a number of classes between one and eight. Also it shows the segmentation result for the optimum number of clusters calculated. In this case the algorithm detects 3 clusters as optimum because there are 3 main different chromaticities in the image, which are so distinct from each other.

The results obtained are better than [Moreira&Costa 96] because the VQ learning strategy is used for the SOM (Self Organizing Map) and the quality factor has been improved. As it is shown the clustering is not perfect because colors undetected exist with a small number of pixels (the window) and other similar colors are taken as one (skin, mouth and some part of hair). But this method gives a good estimation of the skin pixels from the image improving the algorithm of [Stiefelhagen et al. 97] where the skin class is computed off-line.

Then the skin cluster has to be obtained between all the clusters. To do this, the distance between the center of the clusters and the master position, which represents the prototype of the human skin color, is calculated and the closest cluster is chosen as the skin class. Experimentally it has been demonstrated that this master position can be used with people from different races and with a wide range of illuminations.

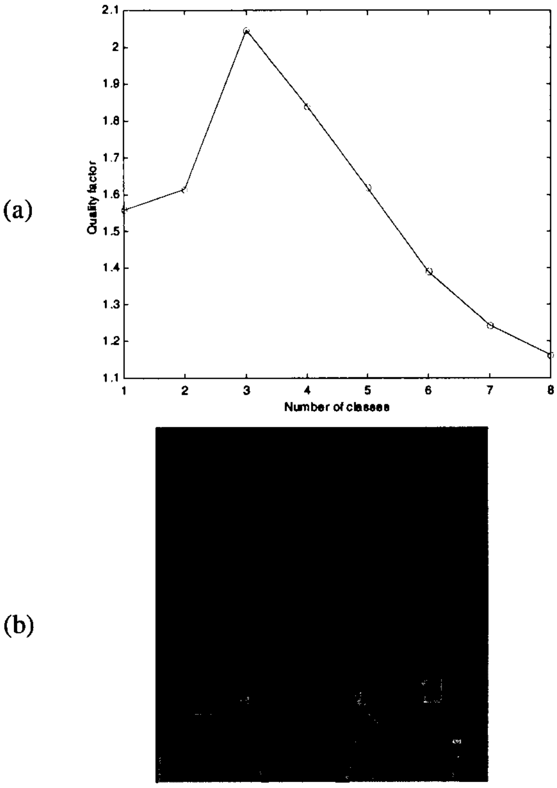

Figure 3. (a) Clustering quality factor (b) Segmented image

4. Skin tracking.

The skin color distribution can be modelled by a 2D Gaussian function, N($\mathbf{m}_S$,C_S), where $\mathbf{m}_S$=(m_{rS},m_{gS}) and C_S are given in the equation (8). These parameters are calculated with the pixels of the skin cluster.

$$m_{rS}=\frac{1}{M_S}\sum_{i=1}^{M_S} r_i \;\; ; \;\; m_{gS}=\frac{1}{M_S}\sum_{i=1}^{M_S} g_i \;\; ; \;\; C_S=\begin{bmatrix}\sigma_r^2 & \sigma_{rg}^2\\ \sigma_{gr}^2 & \sigma_g^2\end{bmatrix} \tag{8}$$

A straightforward way to locate a face is to match the skin color model with the input image to find the face color cluster. Each pixel of the original image is converted into the color space "rg"and then compared with the distribution of the skin-color model. The skin Gaussian function gives the likelihood that a pixel belongs to the skin distribution (Equation (9)). If this value is higher than a threshold (Th) it can be assumed that the color of that pixel is skin.

$$f(x/skin)=\frac{1}{2\pi C_S^{0.5}}e^{-0.5(x-m_S)^T C_S^{-1}(x-m_S)} \tag{9}$$

Most color-based systems are sensitive to changes in the scene. Even under the same lighting conditions, background colors such as colored clothes may influence skin-color appearance. Furthermore, if a person is moving, the apparent skin-color changes as the person's position relative to camera or light changes. In order to use this method in a large range we have used an adaptive skin-color model. The basic idea is to use a linear combination of the known parameters to predict the new ones.

The mean vector $\mathbf{m}_S$ will be a linear combination of the previous average vectors (Equation (10))

$$me=\sum_{l=1}^{v} \alpha_l m_{lS} \tag{10}$$

me is the estimated mean vector; $\mathbf{m}_{lS}$, l=1,...,v are the previous mean vectors; $\mathbf{C}_S$ is the covariance matrix; and α_l, (l=1,...v) are the coefficients for the prediction. If we use the maximum likelihood criterion to find the best set of coefficients for the prediction we obtain:

$$\alpha_j=(\sum_{l=1}^{v} m_{jS}^T C_S^{-1} m_{lS})^{-1} m_{jS}^T C_S^{-1} m_S \quad , \quad j=1,...,v \tag{11}$$

In this case, the covariance matrix $\mathbf{C}_S$ is not estimated because experimentally it has demonstrated that the estimation of this parameter introduces oscillations in the system.

$$C_e = C_S \tag{12}$$

The best result has been obtained maintaining a covariance matrix value constant. To do that we have applied an adaptive threshold proportional to the trace of the covariance matrix.

$$Th = K_{Th} tr\left[C_S\right] \tag{13}$$

The figure 5 shows the evolution of r and g mean and variance for a 50 images sequence. It is also shown the error in the classification of the pixels for the sequence.

It can be seen that the error grows when there is some movement but quickly the system respond adapting its parameters decreasing the error.

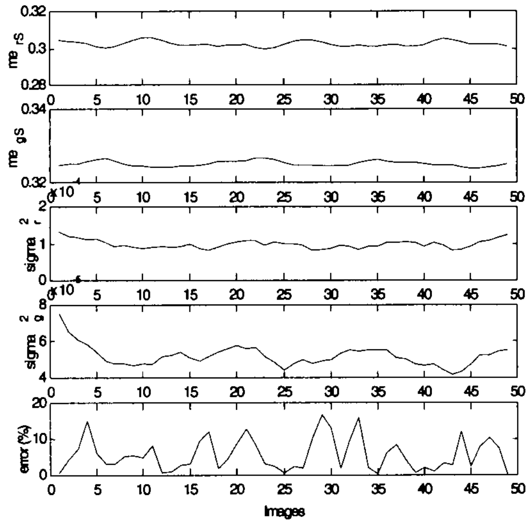

Figure 5. Evolution of the parameters

5. Face tracking

On the skin segmented blob some parameters are calculated: center of gravity (x,y), horizontal (h) and vertical (v) size of the face (Figure 6), being able to obtain the face position and orientation.

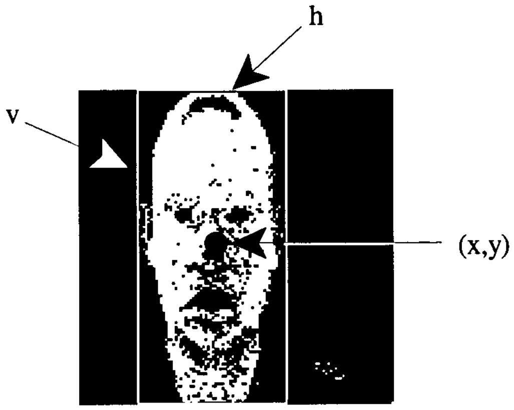

Figure 6. Parameters on the skin blob

The use of estimation theory for tracking and for fusion of information in computer vision is well known [Crowley&Coutaz 95]. A zero-th order Kalman filter is used to estimate two independent state vectors: one to the horizontal variation (X_h =(x,h)) and other to the vertical variation (X_v =(y,v)). Two independent state vectors have been employed because, in our application, the user can do only horizontal and vertical head rotation movements. Then we can take into account that the horizontal and vertical movements are independent, while horizontal size depends on the x center position and the vertical size depends on the y center. The horizontal and the vertical size of the face are estimated as two parameters because the aspect ratio of the face can change with rotations. Both vectors are measured in pixels and each vector is accompanied by a covariance.

The state vectors (X_h,X_v) and theirs covariance matrices (C_{Xh}, C_{Xv}) are estimated in a recursive process composed of three phases: predict, match and update. Then, the estimated state vectors $(\hat{X}_h, \hat{X}_v)$ and theirs derivatives are introduced in a fuzzy classification that codifies the following commands: forward, backward, turn right, turn left, increase speed, decrease speed, stop and no-command. This commands are sent to the low level control by a LonWorks bus.

6. Experimental results

With non-optimized code running on a P200 NT 4.0 the system is able to process 10 images per second. We have tested the skin segmentation clustering algorithm with a face database of 100 people from different races and the system locates the face optimally in the 98 % of the cases. We have analyzed several image sequences in order to test the tracking system and the command generation, obtaining good results.

In figure 7 we can see the results of the identification command in a sequence of 200 images.

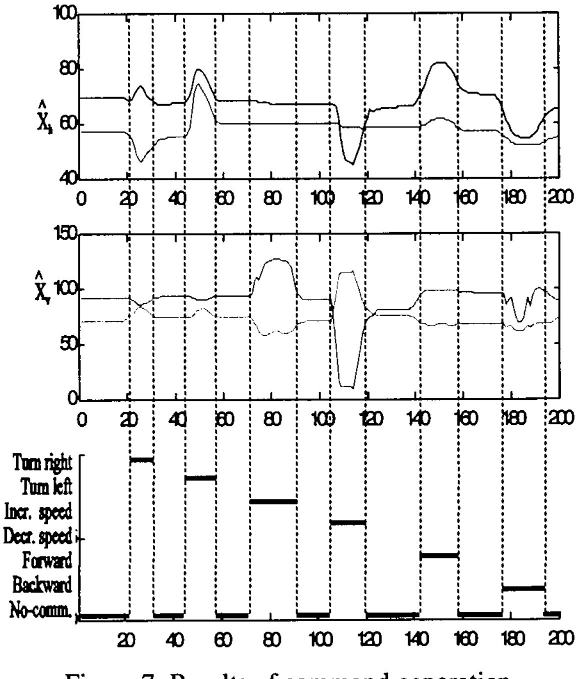

Figure 7. Results of command generation

7. Conclusions and future work

We have designed a real time face tracking using an adaptive skin color model for detecting the skin and a zero-th order kalman filter to estimate position and orientation of the face. It works on complex color images with random background. It is not necessary neither supervised learning nor initial parameters to be introduced. It is able to find faces of different human races at any position, it is adaptive and therefore robust to light and background changes

Now we are working in the tracking of the gaze direction finding the positions of eyes and mouth in the skin area, to estimate its direction using a 3D model of the face.

8. Acknowledgements

This work has been financed by the CICYT (Spanish Interministerial Science and Technology Commission) through the project TER96-1957-C03-01.

9. References

[Canon, 95] Test af Canon UC-X1Hi, HIFI electronik, 1995.

[Crowley&Coutaz 95] Crowley J.L, Coutaz J. "Vision for Man Machine Interaction" EHCI'95. Grand Targhee, August 95.

[Kohonen, 97]T. Kohonen. Self Organizing Maps, Springer-Verlag. Berlin, 1997.

[Mazo et al. 98] M. Mazo, F. J. Rodríguez, J. Ureña, J. C. García, J. L. Lázaro, R. García."Integral System for Assisted Mobility". 2nd International Workshop on Intelligent Control (IC´98)JCIS'98 Proceedings. Editor: Paul P. Wang. pp: 361-364. Durham (USA), 1998.

[Meier et al., 97] R. Meier, R. Stiefelhagen and J. Yang "A preprocessing of visual speech under real word conditions". Proceedings of European Tutorial & Research Work Shop on Audio-Visual Speach Proccesing. 1997

[Moreira&Costa 96] Moreira J., L. Da Fontoura Costa. "Neural-based color image segmentation and classification using self-organizing maps". IX SIBGRAPI, October 1996.

[Stiefelhagen et al. 97a] Stiefelhagen, J. Yang and A. Waibel "A model based gaze tracking system". IEEE International Joint Symposia on Intelligence and Systems-Image, Speech&Natural Language Systems. 1997

[Baluja&Pomerleau 93] Baluja and D. Pomerleau. "Non-Intrusive Gaze Tracking Using Artificial Neural Networks", In Advances in Neural Information Processing Systems, Vol. 6, Morgan Kaufmann, 1993.

[Littmann& Ritter 97] Enno Littmann and Helge Ritter. "Adaptive color segmentation. A comparison of neural and statistical methods". IEEE Transactions of neural networks, Vol 8, N° 1, January 1997.

[Sun&Poggio, 98] Kah-Kay Sun and Tomaso Poggio. "Example-based learning for view-based human face detection". IEEE Transactions on Pattern analysis and machine intelligence, Vol 220, n°1 , January 1998.

[Yuille et al. 92] Yuille A.L., Hallinan P.W. and Cohen D.S. " Feature extraction from faces using deformable templates." International Journal of Computer Vision, pp. 99-111, 1992.

[De Silva et al. 95] De Silva, K. Aizawa, M. Hatori. "Detection and tracking of facial features by using edge pixel counting and deformable circular template matching. IEICE TRANS INF&SYST. VOL.E78-D, No 9, pp.1195-1207, September 1995.

[Heinzmann&Zelinsky 97] Heinzmann and A. Zelinsky, "Robust real-time face tracking and gesture recognition". Proccesings of IJCAI´97, International Joint Conference on Artificial Intelligence, August 1997.

[Yow&Cipolla 95] Yow and R. Cipolla. "Towards an automatic human face localization system." Department of Engineering. In Proc. British Machine Vision Conference, volumen 2, pp. 701, Birmingham, October 1995. Springer-Verlag.

[Yang&Waibel 96] Yang and A. Waibel "A real-time face tracker". Proccedings of WACV'96, Sarasota, Florida, USA. 1996