Before characterizing faces *

M. Castrillón Santana, J. Lorenzo Navarro, M. Hernández Tejera, J. Cabrera Gámez
Grupo de Inteligencia Artificial y Sistemas
Edificio de Informática y Matemáticas
Universidad de Las Palmas de Gran Canaria, 35017 Las Palmas - SPAIN
modesto@dis.ulpgc.es

Abstract

This paper describes the approach for face detection and selection of frontal views, for further processing. This approach based on color detection and symmetry operator application, is integrated in an Active Vision System offering promising results just making use of some opportunistic skills.

Keywords: Face detection, tracking, active vision, feature detection, HCI.

1 Introduction

Current society is characterized by an increasing integration of computers in daily life, human computer interaction (HCI) is currently based on the use of certain devices that are not natural for humans. HCI tools evolution have been notorious and not trivial since its beginning, unfortunately, today accessing to this communication or interaction tools with computers requires training due to the process is not natural as we have already mentioned, and even more it can produce stress. A new generation of on intrusive interfaces development using perceptual capabilities similar to humans is a current trend that focus a great effort due to for the great market that can appear in a short term [1].

Humans are social beings, humans make use of their motion and sensing possibilities for communicating with their environment. In human interaction it is clear the main influence of oral communication, but we should not forget visual information such as body communication, gestures, facial expression which are used simultaneously with sound produced by our throat, but those cues are used also when there is no sound. We can wonder the reason, but it is evident that visual information improves communication and makes it more natural. Facial expressions, the body movements and gestures are natural abilities which are not strange for humans even

Work partially funded by *Spanish Government and European Union Project 1FD1997-1580-C02-02 PI1999/153 and Canary Islands Autonomous Government PI2000/048 research projects*

Figure 1: DESEO tracking sequence

when there are different 'vocabularies' in different cultures. Could a computer make use of this visual information? If HCI gets closer to this schema of human to human communication, accessing these artificial devices could be wider, easier and they could improve even more its presence as human assistants, making human computer interaction non-intrusive and more natural for humans [1].

Many disciplines are involved to solve such a big shuttle. The task that we face would be the integration on an Active Vision System [2] of a module for face detection, characterization and/or description without forgetting people identification and facial and body gestures analysis, using just visual information. In this paper we describe the module for face detection and normalization integrated in a platform designed for general purpose artificial vision applications such as DESEO [3].

2 DESEO overview

DESEO (acronym of Detection and Tracking of Objects in Spanish) is based on a modular conceptual architecture and makes use of active vision for the detection and tracking of mobile objects. DESEO has as low level task reactive system which has the ability to track a moving object, its modular nature eases the mapping of complex activities over a network of modules that implement more simple behaviours. A system developed over DESEO would be able to combine different kind of tasks which are associated to different levels of abstraction. Adapting the general problem, DESEO would behave in an environment where more that one person can catch the attention of a tracking system. DESEO is able to detect different people in an environment using motion detection and/or skin color. Also DESEO is capable of tracking more than one person even when they are not sharing the same field of view in real time using an extended correlation algorithm [3] for the pattern search. Data achieved would be provided to other possible modules that would allow the system different abilities such as recognition, description and so on which have no real time restriction.

In our problem, we have considered the necessity of tracking the head of the person just due to most recognition/interpretation techniques applied to humans make use of the face features as we will describe briefly below. DESEO presents some added features to the ability of tracking different people at the same time. The system incorporates a movement estimation technique for improving its performance when searching the target pattern.

3 Face detection

DESEO can provide visual data to higher level modules which are coordinated in an asynchronous fashion for further processing. In our context input data received by those modules consists of a rectangular window that contains a blob which is likely a human face, next step should be confirming that we have a face and locate its features, which is indeed a huge problem by itself.

The candidate area should be segmented from the background, and should be confirmed as a face in a robust manner for illumination, scale and orientation changes. Once a face has been detected the system would be able to transform properly the image to the work format, as the face can be in any position, with any size, pose or orientation. But first we will pay some attention on the face detection process which is not a simple problem. How do we know that we are in front of a face?

Different approaches has been proposed to solve this problem: pattern recognition techniques, using templates, or models to adapt to face gestures, neural nets and so on. These techniques perform an exhaustive search on the image, while color detection allows to avoid that search across the whole image. For that reason several methods of face detection are commonly based on skin color [4], or combining with other cues as for example edges. Using color and its fast processing feature has allowed to work on this problem using economic hardware. Color also offers a great robustness against scale changes, rotations and partial occlusion when using color spaces where intensity can be excluded for a better modeling of skin color. Too many advantages to avoid the temptation of testing face detection techniques that can be fast enough to be real-time. But also the great amount of techniques evidences that color is not robust under any circumstance, color perception can vary substantially for different environments (indoor, outdoor) specially when the scenery is not under control.

3.1 The approach

A higher level module can asynchronously perform a request to DESEO in order to get the current view tracked as mentioned above. Before applying any further processing we have to confirm the image contains a face in a frontal view. For that purpose the position of the potential face should verify some geometrical constrains and its shape should be similar to an ellipse. Currently the system performs: 1) person detection, using a skin color average value on YUV color space for detecting areas that are face candidates, 2) ellipse approximation, in [5] it is described a technique for obtaining the best ellipse that fits a skin color blob. The dimensions of this ellipse, i.e. the axis ratio, provide an idea of what is not a frontal face, so they can be used as a first filter for removing some blob candidates; 3)face orientation, fitting the ellipse also provides an orientation for the face, for this problem we have considered that a face can be rotated from its vertical position no more than 90 degrees, i.e., the hair is always over the chin, or exceptionally at the same level. The orientation obtained is employed for rotating the source image in order to get a face image where both eyes lie on a horizontal line.

4 Normalization

Commonly recognition systems are trained using a small set of faces from one individual with a known pose. Even when the problem is restricted to frontal views, there could be some bidimensional translations and/or rotations [6] that could affect to the face processing task. That is the reason due to it is necessary to perform a normalization before applying any recognition or interpretation technique. This transformation would allow to avoid differences that are not due to the individuals but the image taken; in [6, 7] training images suffer the same normalization process that will suffer later any new face processed by the system. The transformation is given by the position of some key points on the faces such as eyes, nose and mouth. This procedure would adjust the image face to a standard size while confirming we have acquired a frontal face.

Automatic feature detection techniques have been treated extensively, using static templates, snakes, eigenvectors, the known symmetry operator, performing projections for detecting key points, using preprocessing based on neural nets, SVM, Gabor filters or morphological operators.

4.1 Symmetry detector

Humans facial features offer a symmetry appearance, for example our eyes are rounded by an elliptical shape, in [6] it is considered the use of a symmetry operator in order to locate interesting elements on the face without an a priori knowledge on their shape. This operator presents a high computational cost that can be reduced using some heuristics.

The symmetry transform as defined in [6] works over an edge map obtained using a local operator which is calculated using fast and rude detector as the performance of the symmetry transform is not affected by the use of a fine edge detector. This measure takes into account the distance and orientation of edges around a point. The symmetry measure is the addition of all the contributions provided by pair of points located under a threshold range around a concrete point using a weighted gaussian.

4.2 Automatic normalization

Ellipse orientation can be used to adjust the face to a vertical position before locating eyes which in our experiments has been the points that has provided a better response to the symmetry operator. Ellipse center can also be used for bounding an area where applying the symmetry operator in order to reduce the computational cost and even more we could restrict to calculate the operator in those areas where skin color is not detected, as happens commonly with our eyes even when they are closed. Once we have a symmetry image over a concrete image area which has been selected using the color cue and the ellipse that best fits the blob image. Searching for maximal values in restricted areas of the symmetry image we can obtain a couple of candidate eyes.

Once these potential eyes have been located the system proceeds performing some tests based on contextual knowledge in order to validate that the eye positions in the ellipse are coherent for a frontal view with ellipse orientation and their inter-distance is also coherent with ellipse dimensions. After these tests, that distance is used for scaling the image to a standard size while other approach would consider three key points (eyes and mouth) and perform a morphing of the original image to the standard one.

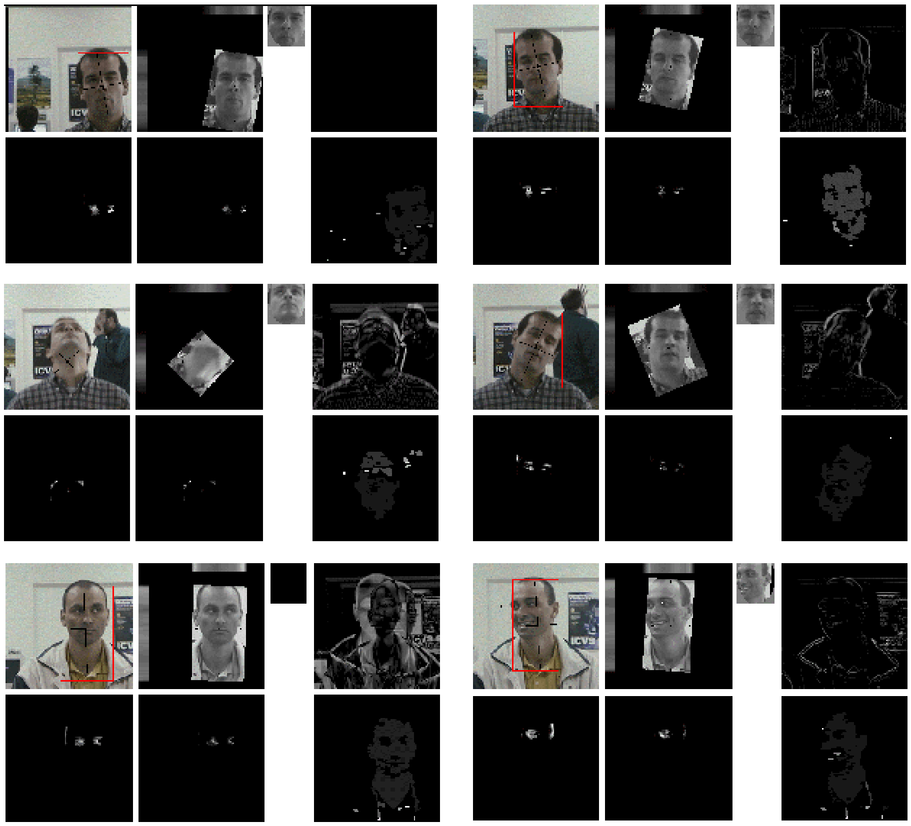

Figure 2: Face Detection and Normalization samples

5 Experiments

5.1 Detection experiments

Experiments have been performed for different sequences, each one containing 1000 frames taken under similar lighting conditions, where the subject is sat in front of the robot head moving freely but without running away from the chair. In figure 2 some processed frames are presented with the symmetry images, difference images, the blob images and the normalized images when they were obtained. Typical processing time for a blob size of 60x85 using a Celeron 433 Mhz is in the range 100-200 msegs depending on the area detected as no blob. The table presents results of correctly automatically detected frontal and lateral views against the number of them labeled manually (Automatic/Manual), as well as images rejected and false positive.

Sequence	Frontal	Lateral	Refused	False positive
antonioseq	380/555	300/375	320	0
antonioseq2	220/614	240/285	540	0
cayeseq	73/474	380/520	540	0
cayeseq2	38/539	365/455	540	0
modestoseq	290/615	275/337	420	5
oscarseq	520/770	90/95	380	2
danielseq	485/830	110/160	400	2

As can be observed in the table, the system do not provide the same result to any sequence, it is certainly due to the subject, as some subjects skin is better detected or te physonomy of the eye area is different among them (for example with cayeseq, left eye is normally bad detected due to the combination of light and his eye area).

6 Conclusions and Future work

6.1 Color detection improvement

Color detection affects system performance as many faces labeled by humans are frontal are not processed by the system due to the skin color blob dos not seem to be correct. The great variety of color spaces for detection seems to prove the impossibility of achieving a general method just using a simple approach as we do. As color can be a valid an interesting clue, some color spaces has been examined in order to check which one would offer for our environment a better performance. For that purpose it has been used a feature selection approach providing some results that has been used later for training a classifier instead of just using the previous approach that was based on a set of rectangular areas on YUV space.

For that purpose some different color spaces has been studied for a set of sequences acquired with DESEO, this color samples contains 31752 samples corresponding 17483 to no skin samples and the rest to skin samples. Color spaces used were YUV, normalized red and green ($\tilde{r}$,$\tilde{g}$), a perceptual uniform color system, RGB and $I_1 I_2 I_3$

GD [8] measure provided a features sort as follows: $\tilde{g}$, I_3, V_f, V, I_2, $\tilde{r}$, U_f, U, R, I_1, Y, Y_f, G and B. For color detection improvement next step would be training a classifier that could be a RBF as a region of similar color in the scene maps to an elliptical region in color space.

6.2 Robustness

Another improvement that is being considered is the fact that we should keep in mind that we are processing along time so temporal coherence could be considered in order to improve face detection. As faces are taken constantly we could consider the information from the previous frame in order to validate the next one. On each frame we could take information about the biggest blob such as ellipse orientation, its position and dimensions, and for example we could give a kind of persistence to each pixel that was considered skin in a previous frame.

References

[1] Alex Pentland. Looking at people:sensing for ubiquitous and wearable computing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, January 2000.

[2] Y. Aloimonos, I. Weiss, and A. Bandyopadhyay. Active vision. *Int. Journal of Computer Vision*, 1(4), January 1988.

[3] M. Hernández, J. Cabrera, M. Castrillón, A .C Domínguez, C. Guerra, D. Hernández, and J. Isern. Deseo: An active vision system for detection, tracking and recognition. *Lecture Notes on Computer Science, 1542*, 1999. Springer-Verlag, ICVS'99, Gran Canaria.

[4] Yogesh Raja, Stephen J. McKenna, and Shaogang Gong. Tracking and segmenting people in varying lighting conditions using colour. *Proc. of FG'98*, 1998.

[5] Karin Sobottka and Ioannis Pitas. A novel method for automatic face segmentation, face feature extraction and tracking. *Signal Processing: Image Communication*, 12(3), 1998.

[6] Daniel Reisfeld and Yehezkel Yeshurun. Preprocessing of face images: Detection of features and pose normalization. *Computer Vision and Image Understanding*, 71(3), September 1998.

[7] C. Wong, D. Kortenkamp, and M. Speich. A mobile robot that recognizes people. *IEEE International Conference on Tools and Artificial intelligence*, 1995.

[8] J. Lorenzo, M. Hernández, and J. Méndez. Gd: A measure based on information theory for attribute selection. *Lectures Notes in Artificial Intelligence*, 1484:124–135, 1998.