CONTRIBUTIONS to SCIENCE, **1** (2): 159-173 (1999)
Institut d'Estudis Catalans, Barcelona

Possibilities of man-machine interaction through the perception of human gestures

Josep Amat[1], Alícia Casals[2], Manel Frigola*[2] and Jaume Pagès[2]

1 Institut de Robòtica Industrial (UPC/CSIC), Barcelona

2 Departament de Control Automatic. Universitat Politècnica de Catalunya

Abstract

As man-machine interaction grows there is an increasing need for friendly interfaces. Human-machine oral communication as a means of natural language interaction is becoming quite common. Interpretation of human gestures can, in some applications, complement such communication. This article describes an interpretation of gestures procedure. The system is based on a computer vision system for the detection and tracking of a human operator and the interpretation of a specific set of human gestures in real time.

Resum

A mesura que les màquines s'utilitzen interaccionant cada cop més amb les persones, la necessitat d'interfícies més amigables esdevé una necessitat creixent. La comunicació oral persona-màquina com una forma d'interacció utilitzant el llenguatge natural és cada vegada més usual. La interpretació dels gestos humans pot, en certes aplicacions, complementar aquesta comunicació oral. Aquest article descriu un sistema d'interpretació dels gestos basat en la visió per computador. El procés d'interpretació realitza la detecció i seguiment d'un operador humà, i a partir dels seus moviments interpreta un conjunt específic d'ordres gestuals, en temps real.

Keywords: Control by gestures, gesticular commands, man-machine interface, human body tracking.

Man-Machine Interface (MMI) is crucial for optimising interaction between humans and computers. Humans are quite efficient in dealing with intuitive, logical, qualitative and conceptual problems, while computers are more suitable for arithmetic, boolean logic, quantitative and algorithmic operations. Moreover, computers are characterized by large reliable data memories, high computation speed, etc. In spite of these traits, artificial intelligence is still very limited making human-computer cooperation necessary. Therefore human-machine communication and interaction is a key factor in the development of efficient semi-automatic or computer assisted control systems.

Besides classical keyboards, push-buttons, mouses and graphical interfaces, the need to incorporate natural language as an additional element has led to the development of speech recognition and voice synthesizing systems. Voice recognition systems facilitate the development of oral control systems that are often more flexible than classical interfaces, or can transmit commands at a higher speed.

Oral computer communication can be complemented by other natural communication based on interpretation of gestures. These two means of natural communication complement each other in clarifying ambiguous expressions. Oral expressions such as *this* or *that* can reinforce some arm gestures or, alternatively, gestures like pointing at a given object can show a desired target which is complementary to oral commands. Gesture interpretation constitutes a new means of man-machine communication. Human gestures can provide the computer with complementary and relevant information which is more versatile, flexible and intelligent than other communication forms.

A significant amount of work has already been done in this relatively new research field. This study presents developments on the detection of a human body in a setting, the tracking of movements and the interpretation of gestures. This project has already given very reliable results. The gesture interpretation system operates at high speed and low cost.

This study aims to address the automatic interpretation of clear, expressive gesture commands. These gestures are

* Author for correspondence: Manel Frigola, Departament de Control Automatic. Universitat Politècnica de Catalunya. Pau Gargallo, 5, 08028 Barcelona, Catalonia (Spain). Tel. 34934016971, Fax: 34934017045. Email: frigola@esaii.upc.es

carried out with either one or two hands, as is common in human behaviour, in conditions in which oral commands are not possible. A common example is the gesticular indications of a driver's assistant helping the driver during a maneuver. The assistant gives instructions such as *forwards* or *backwards, to the right* or *stop*. New possibilities are now being discussed about the detection of strange human behavior [42] through abnormal postures and gestures. Actions such as a thief trying to enter a car, a person planning to commit suicide by jumping in front of a train or threatening someone with a knife or gun, can provide clues for their interpretations. Developments in interpretation of postures and gestures can open up this new application sector.

This study aims to provide an automated control system with some kind of gesture command inputs. Such commands can be either complementary or occasional corrective commands to be incorporated into a pre-established control program. The aim is also to design a system whereby these orders could be given by any operator without the need to wear special clothes or marks for the task, as happens, for instance, when helping a pilot to position the plane correctly in front of a finger in airports.

Thus, the developed system operates from the analysis of the images obtained from the setting, by tracking an operator who is not specially dressed for the task or equipped with specific identifying elements. The tracking starts by automatically detecting the operator's position in the setting thus getting the three-dimensional position of his arms. It is also possible to restrict the active search area to a predefined area of the setting to reduce the error probability by avoiding the detection of unwanted moving objects and to decrease the search time. The system then interprets the operator's postures, movements and gestures that are significant enough to be interpreted as control commands.

Previous work

The first studies on computer vision for the recognition of human body movements were performed by J.O'Rourke and N.I. Badler [43]. They used synthetic images composed of rigid segments and joints to model a body. The simplicity of such images, which can be perfectly segmented, enabled the study of the constraints necessary to track several kinds of human movements. Later on, D. Hogg [23] started to use natural images without marks. This work was limited to the tracking of only very primary gestures. The works by M. Yamamoto and K. Koshikawa [55] and some years later by M. Dhome, A. Yassine and J. Lavest [18] formally extended the study to more complex structures by applying the kinematic and dynamic models used in robotics.

Working in parallel, other researchers developed systems to track body movements using devices in contact with the user's body to facilitate the detection of body parts, for instance LED diodes or reflectors located in the articulations [6, 41, 49]. Other systems are based either on magnetic position sensors, or on exoskeletons [52] requiring the operator to wear a mechanical structure like the Hardyman exoskeleton built in the sixties by General Electric. Myoelectric sensors that convert the muscle movement into signals, from which it is possible to detect the operator movements [1], or master-slave devices based on commercial data gloves constitute more sophisticated devices. The use of hybrid techniques [29] using both magnetic sensors and cameras detecting some marks (fiducials or landmarks) give improved performances since they combine the robustness of magnetic sensors with the precision of computer vision.

The use of computer vision as a remote sensor of human movements facilitates operation in natural environments, without the need for invasive techniques and consequently offers more possibilities for practical applications [2]. Several computer vision techniques are used to detect and track the movements of the human joints: speed estimation through optical flow and finite element methods [24, 37]; use of deformable functions [32] driven by 3-D points and 2-D edge measurements; or comparison with a model of contours and regions extracted as targets [20,30,39]. Most of the above mentioned studies, as well as others, are based on the assumption that the initial configuration of the articulated body is known. This requirement is fulfilled either with manual intervention or by forcing the user to adopt a determined initial posture. From this situation the joints and model size are adjusted and the system operation is carried out incrementally, dynamically updating every new position of the body part from the analyses of differential movements.

Sport constitutes an area where gesture analysis can provide significant information. The exercises of elite athletes have been evaluated using articulated models of the human skeleton. Assessments have been performed from the point of view of their performance, and even from the beauty in their execution. Posture and gesture models are adjusted image by image either manually or semi-manually. In the latter, the system interpolates intermediate postures and the operator is only required to adjust some key images [57].

Since common tracking techniques are not able to deal with sudden or quick movements or with occlusions, they are not robust enough to track all kinds of movement. To solve this problem other methods focus on the detection of certain body parts rather than tracking. Characterization of body parts can follow three approaches: pattern matching, adjustable models, and singular shapes of interest.

In the first approach the body parts are represented by partial reference images (head, arms, body and legs). The image pattern can either be gray level images of some main features [8, 11, 12] or constitute more compact representations [38, 50]. These methods work efficiently provided the images are taken from a predefined point of view, which assures a good fit with the model. Research is now focussing on the ability to transform the acquired images so as to be independent of the original point of view.

In the methods based on parameterized models, these models are usually adjusted to the parts present in the image by minimizing an energy function [21, 25, 28, 44, 56]. The adjustment, by means of deformable curvature shapes, continues as the human body adopts different postures. The model parts are softened shapes of the preprocessed image.

The third approach is based on the search in one or several images of singular shapes related to body parts. Most of these methods work from the same stages: segmentation, contour extraction and object recognition, among others. Image segmentation enables to reduce the amount of data to process. Segmentation techniques based on the skin color [26, 31, 54], thermal images [35], comparison between images [4, 7] or movement information [2, 10, 33, 40] are the most commonly used. Feature extraction is frequently based on the localization of silhouette high curvature points [22, 35, 51]. The order and position in which these curvatures are located provides information about the posture [22]. The use of additional information such as color or texture [5] makes the process more robust.

In the research field of dynamic gesture recognition, attention is usually focussed on the recognition of facial expressions [34, 56, 46], head movements [53] and hand gestures [45, 54]. Whatever the body part of interest, human movements or postures are usually characterized by patterns which represent their spatial displacement and temporal motion. In order to adapt the patterns to differences in velocity and position, probabilistic space-temporal matching techniques, such as dynamic time warping [16] or Hidden Markov Models [45] are used.

System description

The first step in MMI through gesture is the detection and tracking of the human body. Human movements are detected by estimating the position of body parts over time The estimated position sequences describe the body movement and, from this description, it is possible to interpret the gestures the person wishes to express. The whole process, whose functional structure is shown in Fig. 1, can be briefly described by naming the main processing modules:

- Low level processing for movement detection
- Feature extraction and singular point detection
- Singular point classification
- 3D position measurement and posture data validation
- Temporal filtering of the estimated body location and postures
- Human body position and posture prediction
- Tracking of singular points and dynamic search window generation
- Gesture recognition

The system operates based on the information provided by a model of the human body, which is specifically designed to recognize a shape of human body. Several singular points are detected from some extracted features of an image. A first classification of these points is made of the most significant body parts, the head and arms. The search

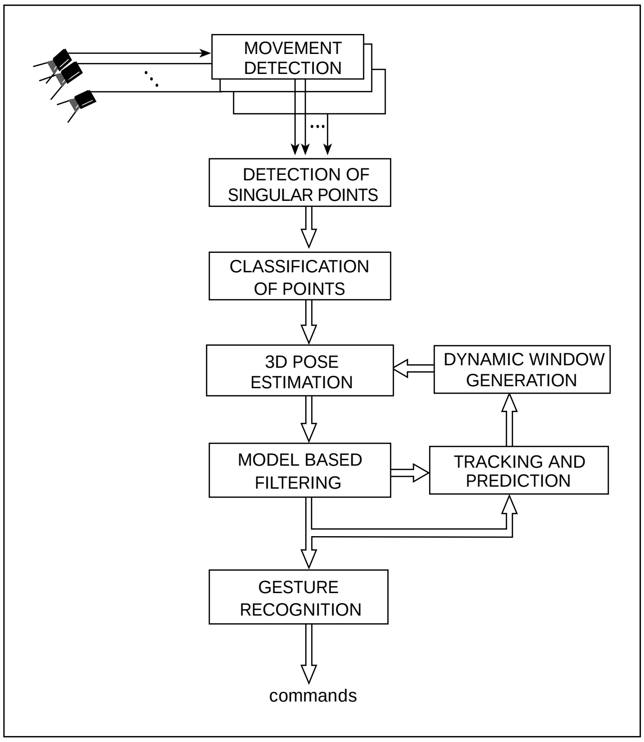


Figure 1. Functional system structure

of singular points and their classification to identify these relevant parts is also based on the prediction and tracking of the trajectories of each part. The tracking algorithm is suitable to compensate the possible missing detections of the recognition system provided they are short. Missing detections are due to occlusions (total or partial occlusion of a body part) or to the lack of visibility caused by a low contrast between the body and background. The recognition system can provide new position references of the body parts when they appear clearly in the image for the first time or after a short occlusion period.

Background extraction

Image segmentation is one of the main problems to tackle in computer vision. The segmentation process consists of a set of image operations or transformations that aim to detect the parts of the image that are of interest for a given application. Image segmentation significantly reduces the amount of data to analyze in later images.

The selection of the adequate segmentation technique greatly depends on the kind of setting to analyze and its environmental conditions. In very simple images, those where the object or objects to be extracted from the background present a high contrast with respect to the remaining parts contained in the image, the process of image thresholding is good enough. In this kind of segmentation the only key factor is the possibility to find an adequate binarization threshold. In other cases, other image features such as the object contours, color or texture are used as object or region discriminating factors. In images where some kind of homogeneity characterizes the image of the object, region growing techniques can be adequate. These techniques consist of selecting one or more small areas in the image, the seeds, and build from them, spreading around the neighboring pixels while they maintain a predefined level of similarity.

When scenes are complex and there is a lack of discriminating features, the above mentioned methods cannot be used to segment the desired objects satisfactorily. However, it is possible to resort to the analysis of temporal variations in image sequences, provided the objects to be segmented are moving. In this case situations there is not a true segmentation but an extraction of the moving objects from the scene background.

The effective use of human figure extraction in natural environments, either indoor or outdoor, relies on segmentation techniques which are not dependent either on the heterogeneity of the possible elements in the setting and its lighting conditions, or on the movements of the person.

Image subtraction is one of the most common movement detection techniques due to its computing simplicity. Nevertheless, this technique is quite sensitive to noise and light fluctuations that prevent the use of threshold values sensitive enough. Therefore image subtraction techniques are used in specific plateaux, under controlled lighting conditions and typically with a high contrast between the moving object and the background. Optical flow is another common technique for movement detection. It is based on the determination of the spatial displacement of each pixel from one image to its correlatives.

Owing to its simplicity, the subtraction method is used in our system, but to improve its performance in complex settings a pixel by pixel comparison is carried out from the estimated gradient vector instead of using only its absolute value. In this case, image subtraction is performed as follows:

$$\left|\vec{G}_t(x,y) - \vec{G}_{t-1}(x,y)\right| \qquad \text{(Eq. 1)}$$

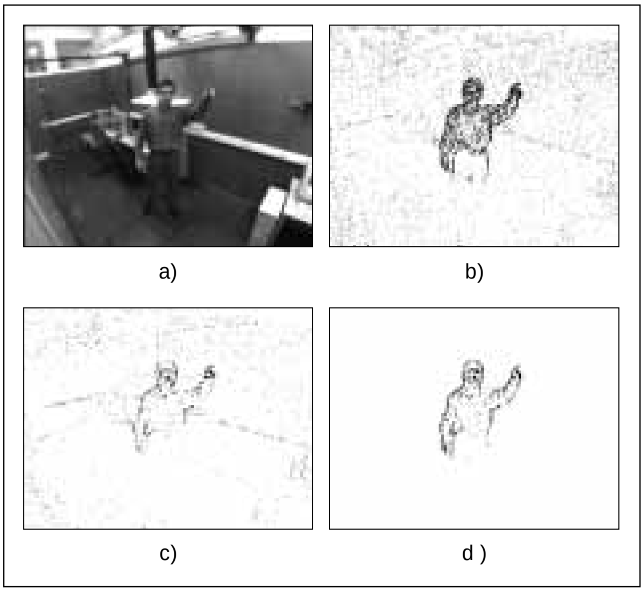


Figure 2. Successive images subtraction. a) One original image, b) gray level difference, c) gradient module difference, d) gradient vector difference

Where $\vec{G}_t(x,y)$ represents the measurement of the gradient vector computed at position *x,y* of the gray level image *I(x,y)*, taken at instant *t*. The advantage of comparing images using the gradient vector instead of the gradient module is the increase from one to two dimensions in pixel description, thus enhancing their characterization. Consequently, the movement detection is less ambiguous and the neighboring pixels within the operating window contribute to increasing the sensibility of movement detection. This is specially crucial in images with low contrast and low signal to noise ratio.

In spite of this advantage, image gradient subtraction is still sensitive to lighting variations. In natural environments, or in environments with fairly controlled lighting conditions, it is necessary to use a more robust comparison. The following new expression includes the gradient direction which does not depend on variations of lighting intensity:

$$\left|\mathrm{Atan2}\left(\vec{G}_t(x,y)\right) - \mathrm{Atan2}\left(\vec{G}_{t-k}(x,y)\right)\right| \qquad \text{(Eq. 2)}$$

Where *Atan2* is the extension of the arctangent function to two dimensions.

Since equation (1) is simpler than (2) the former is normally used in our system when lighting conditions are fixed, and only when lighting variations could decrease reliability is it necessary to rely on (2). Fig. 2 shows the resulting subtraction of a sequence with significant lighting changes using different methods: from gray level images (a), gray level difference (b), using gradient module (c), and gradient vector (d). Subtraction of gray level images is even more sensitive to lighting conditions since luminance variations affects more pixels of the image.

Moving object contrast, lighting conditions and the setting complexity condition the quality of the detected silhouette. Fig. 3 shows some results of silhouette extraction from motion. It is possible to appreciate the limited segmentation quality. The effects of shadows, reflections, poor contrast and noise hinders the correct silhouette detection thus producing a lack of continuity, some missing body areas and, the detection of unwanted contour segments.

Feature extraction and 3D data obtention

Several methods base image analysis on following human silhouette contours to detect body parts [27, 36]. They need to operate over the body silhouette extracted from its back-

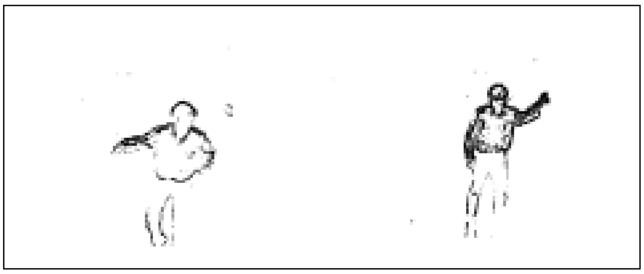


Figure 3. Movement detection with some body segments missing from the setting of Fig. 2

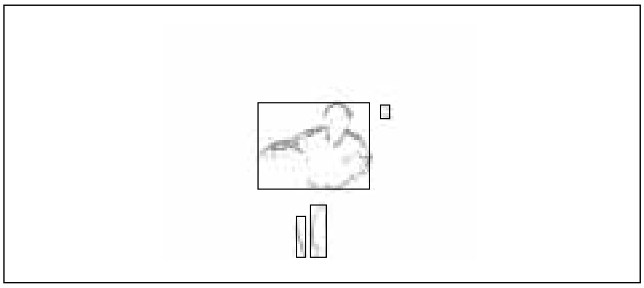


Figure 4. Pixel clustering

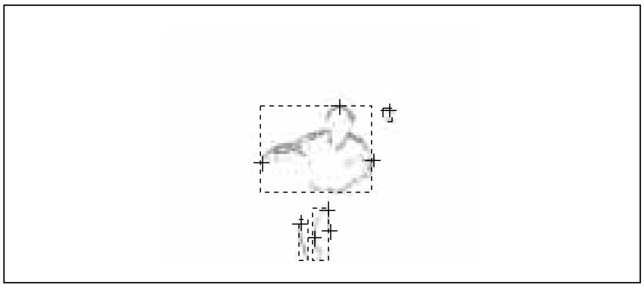


Figure 5. Candidate singular points

ground. Others use identifiable natural marks on the body or clothes surface. The identified marks are taken as patterns that are used to find their homologues in the following images.

The adaptability of the method to different kind of environments, without imposing strong operating restrictions, is achieved by splitting the problem into three steps:

- Extraction features
- Detection of singular points
- Classification of singular points

The features selected to detect and locate singular points are the clusters of pixels that verify some pre-established heuristic conditions. First, the clusters considered are those whose distance between pixels is less than a maximum value, this value being chosen according to a compromise between efficiency and computing cost. A second parameter is the cluster size, those clusters that do not reach a predefined perimeter length are rejected. These data clusterings define the areas where there is a significant movement. Fig. 4 shows the clusters detected in an image.

The next step is the detection of singular points of the body from these clusters. The singular points are extracted from the pixels on the rectangle boundaries as well as from those having absolute and relative maximum ordinates and also their relative minimum (Fig. 5). The absolute minimum has not been considered since it is usually meaningless due to ground reflections. These considerations are based on the assumption that the operator is standing or seated, and consequently the direction of exploration is from the head to foot. The radii of the analysis window is fixed using metric units, since the size measured in pixels depends on the camera-person distance, and it is computed from the anterior estimation of the person's position.

It is highly probable that the singular points corresponding to the head and arms are extreme points or relative maxima or minima within the rectangles defining the clusters. They are classified depending on their relative position and according to the human body model. First the absolute and relative maxima ordinates of these clusters and also their relative minimum are ordered from highest to lowest by counting down, from the considered singular point, the number of image body lines that cut the vertical axis crossing this point. The considered lines are limited to those verifying that the segment limited by the body boundary surpasses a given threshold. Fig. 6 shows the difference of lines detected from the singular points of the head and hand. The point with an absolute or relative maximum ordinate and from which the maximum number of lines are counted was classified as belonging to the head. This point coincides sufficiently in the two stereo images. Since it is the most prominent point in the corresponding image area, its detection in the two stereo images presents no ambiguity when there are no occlusions.

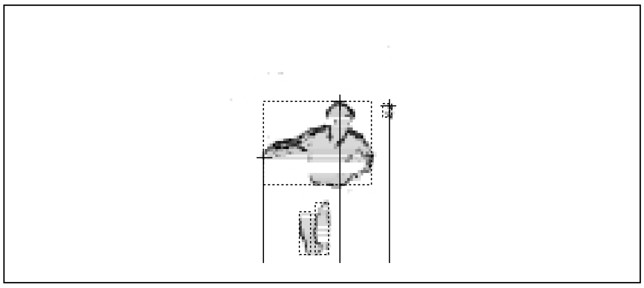


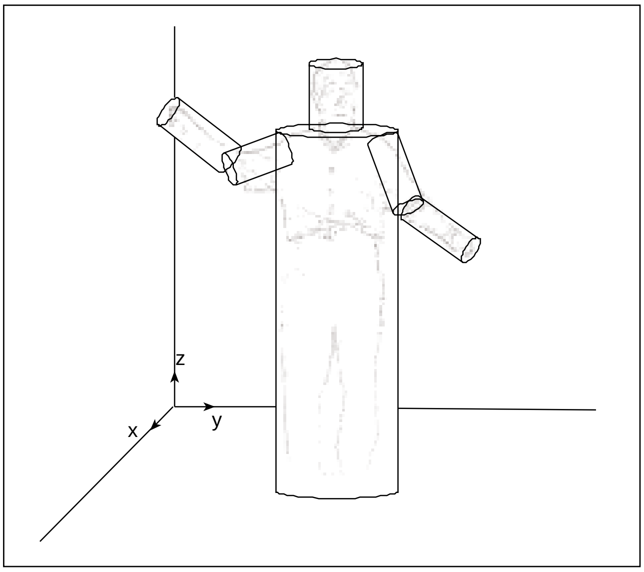
Figure 6. Head and hand classification from horizontal line counting

Model based target validation

With the aim to robustly detect and recognize human shape and posture and to avoid false detections, a human body model is defined to validate the extracted silhouettes. The model was defined based on a compromise between simplicity and speed on one hand and efficiency on the other. It was designed according to the human body structure, its shape and moving capability. The articulated body state is defined by a set of variables that, at a given instant, defines position, speed and acceleration of the different parts that make up the model.

Many of the models proposed in the literature are oriented to represent esthetic movements and gestures [19, 47] or are based on mathematical formalisms [20, 28]. These models have been designed prioritizing their functionality more than trying to reduce algorithm complexity. The aim of this work has been to define a model which is applicable in real time low cost systems but simple enough to guarantee reliable performances. Therefore, the model was designed as an articulated structure composed of geometrical primitives.

The imposition of some anthropomorphic constraints and the availability of some dimensional measures make it possible to reject false detections without the need to apply the model, thus reducing operation time. Therefore, it is possible to reject the shapes that do not fit an adequate profile. The person model adopted is constituted by a set of cylinders that



Fiigure 7. Simplified human polycylindrical model.

fit the moving parts profile. The model consists of two coaxial cylinders that are adjusted to the head and body, and also a set of up to four cylindrical surfaces that are adjusted to the hanging elements, that can correspond to the arms (Fig. 7).

The model is defined by means of several parameters that correspond to the dimensions of the model parts and their tolerances. The parameters are:

- Lower body radius and height (R_T and H_T)
- Head radius and height (R_C and H_C)
- Lower arm radius and length (R_A and L_A)
- Upper arm radius and length (R_B and L_B)

These parameters and their tolerances are defined in metric units providing an adequate operating range to enable the fitting of the model to different kinds of users, Fig. 8 (a). The adequate selection of H_T allows the detection of either seated or standing subjects.

The definition of a model configuration, the posture, is based on some variables that define position and the state of joints as follows:

- The head upper point position, C = (X_C, Y_C, Z_C)
- The angle θ_1, corresponding to the shoulder deviation with respect to the setting reference frame
- The three angles that define each upper arm orientation $(\theta_2, \theta_3, \theta_4)_{Arm\ 1}$ and $(\theta_2, \theta_3, \theta_4)_{Arm\ 2}$
- The elbow flexion for each arm $(\theta_5)_{Arm\ 1}$ and $(\theta_5)_{Arm\ 2}$

Fig. 8 (b) shows the model joint reference frames using the Denavit-Hartenberg representation for articulated kinematic chains.

Each time a moving object is detected the system obtains the three-dimensional coordinates of the prominent points.

The measure of the 3D head position, with respect to the setting reference frame, is computed by triangulation from the points classified as head in two stereo images. The head position is used to center the coaxial axis of the two main cylinders of the model, the head and body. These cylinders are obtained as the silhouette circumscribed volume. The projection of these volumes to the image plane delimits the body parts. Once the different singular points are classified as head and arms, the estimated body position and posture have to be verified. The point considered as the upper limit of the head axis C are identified, as well as the other possible points B_1 and B_2, the system verifies the coherence of the radii R_1 and R_2 corresponding to the points B_1 and B_2 and also their heights: H_c for the head, and H_1 and H_2 for points B_1 and B_2. Fig. 9 shows the model fitting procedure corresponding to the three basic components, the head and the two arms.

The head point is used to center the coaxial axis of the two main cylinders of the model. This first model adjustment carried out by stereovision calculations is considered to fulfill the «person» conditions if the corresponding cylinders verify the tolerances of the model parameters.

The cylinders to fit with different body parts, trunk, head

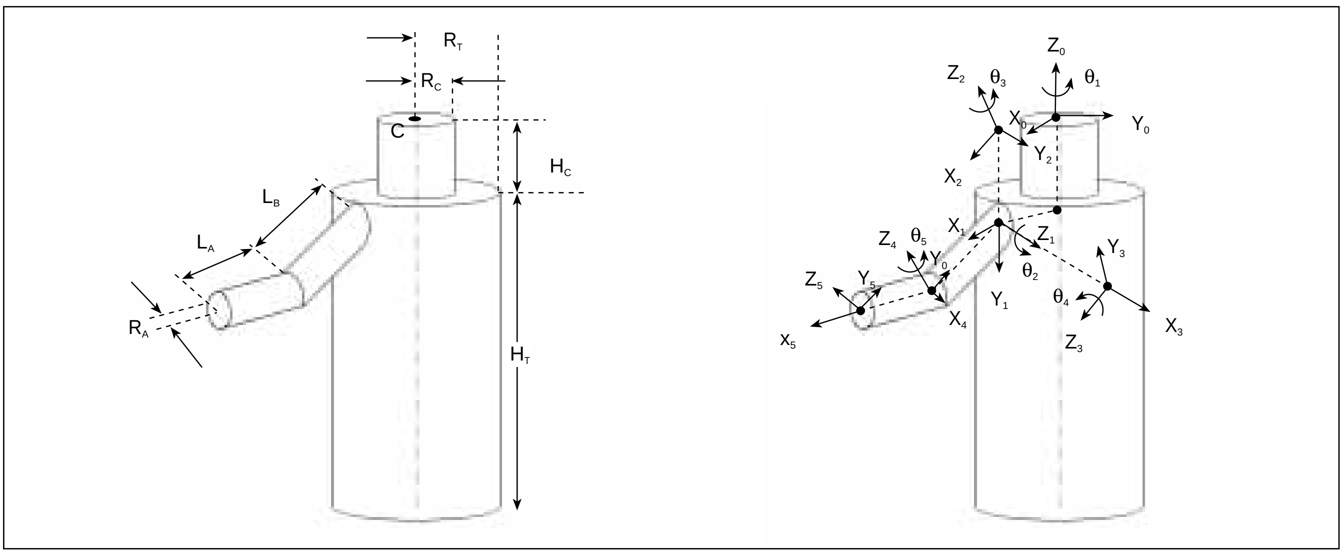
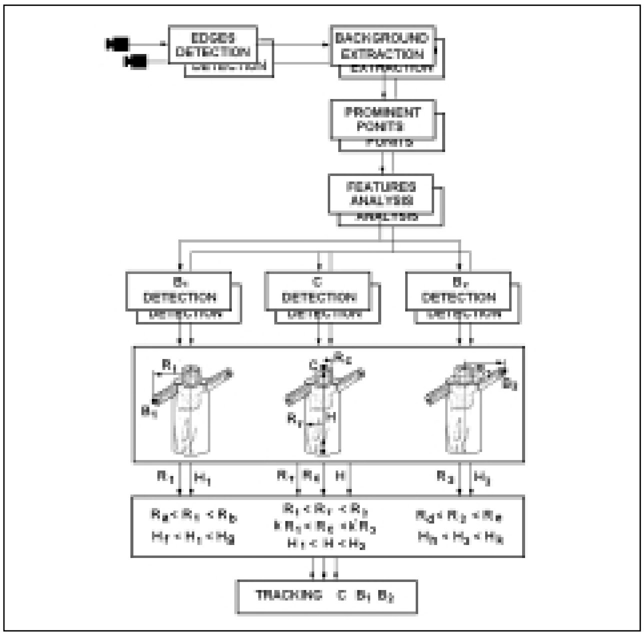


Figure 8. Parameterized model a) Parameters of the model links b) Model joint reference frames



*F*igure *9.* Model fitting procedure

and forearm, as well as the parameters used to verify their correspondence with the predefined model within the accepted tolerance limits are shown in Fig. 10.

Movement tracking

The aim of this step is to provide reliability and to assure that the data supplied by the previous steps are physically feasible. Their validation requires the use of anthropomorphic constraints, the lengths and radii of body parts. Other temporal restrictions can be imposed according to the expected range of speed of the moving parts. For instance, a natural movement can be characterized by its soft trajectory.

$q = (C_x, C_y, C_z, \theta_1, (\theta_2, \theta_3, \theta_4)_{Arm\ 1}, (\theta_2, \theta_3, \theta_4)_{Arm\ 2})$ being the estimation of the model fitted to the body position and posture, at time instant *t*, speed and acceleration are estimated from the first and second derivatives of the ten polynoms that approximate each component of vector *q*. From this continuous estimation the space position of future singular points is predicted. A search window in the image is then generated to restrict the computing area just around the point of interest. The size of this window increases with the sampling period as does the prediction error.

The tracked movement is filtered to smooth the noise produced by sampling. The evolution of speed and acceleration parameters also facilitate the distinction between the movements corresponding to a gesticular command, according to the gesture dynamics, from those due to undesired sudden movements that cannot be interpreted as part of a gesture.

Filtering and interpolation of the sequence of *q* vectors allows the interpretation of a soft gesture movement. The kind of filter and interpolation procedure is selected as a compromise between the speed of foreseen gestures and the softness of estimated movement.

Interpretation of gestures

The ability to perceive particular body motions as meaningful gestures is therefore essential if computer systems are to interact with humans. Some authors have adopted appearance-based techniques for gesture recognition, where the representation of human movements is based on a sequence of pattern images. This representation allows the modeling of complex poses for which no simple 3-D models or recovery methods are available [15, 16, 17]. However, there is no explicit information about the geometrical aspects of gestures. Several applications need to measure the head and hand positions and orientation in order to know what the user is pointing at. Then, the recognized gesture must be complemented with this geometrical information (e.g. orientation of the arm) for its subsequent high level interpretation of human actions.

This MMI based on the tracking of a person's gestures has given satisfactory results in applications where the coded commands transmitted from the human operator to the

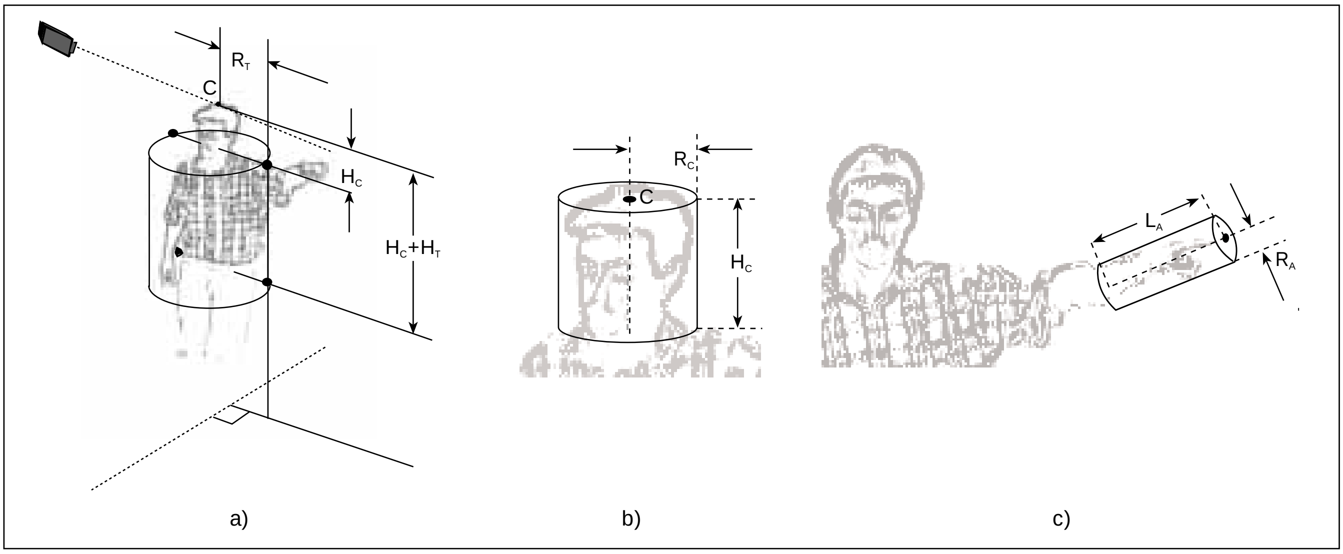


Figure 10. Cylinders and model parameters of several body parts. a) trunk b) head and c) forearm.

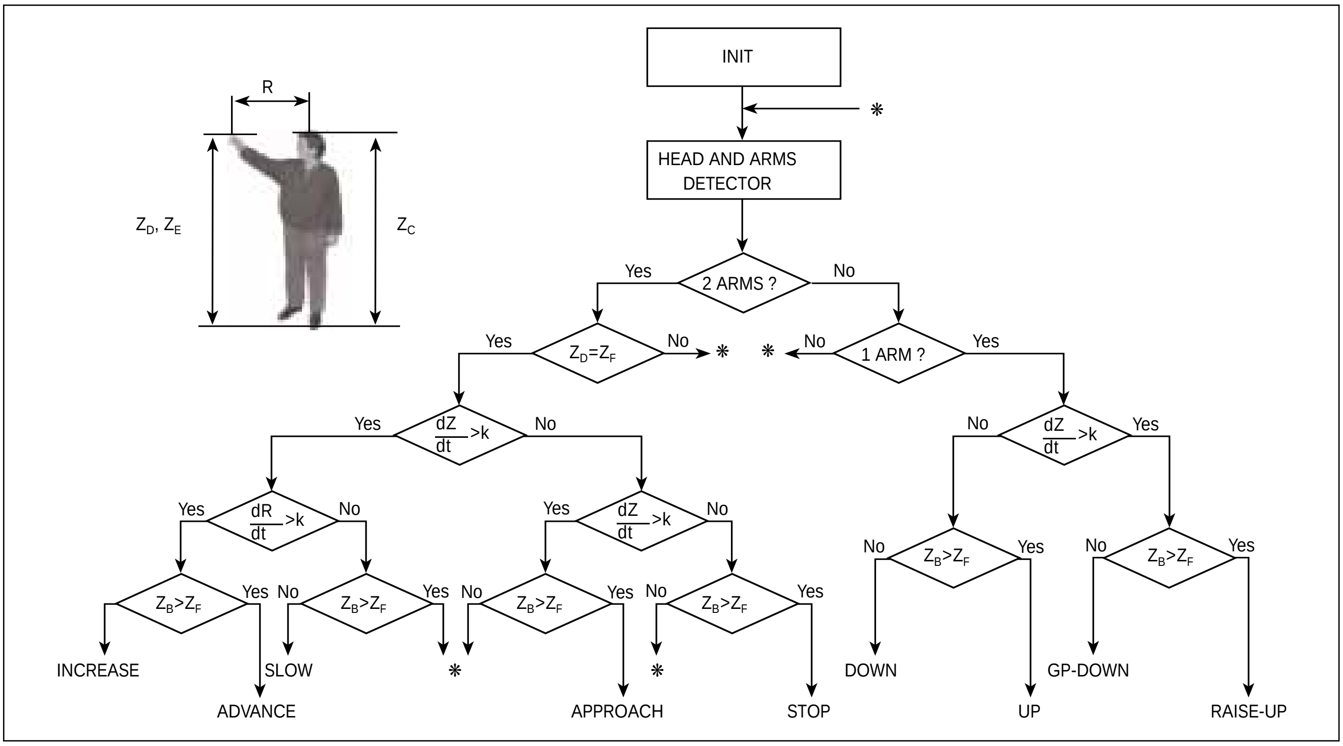


Figure 11. Flow diagram of some gestures interpretation

machine are clear and simple. The interpretation of these simple commands is based on the obtention of a new simplified model of the tracked body, extracted from the skeleton of the polycylindrical body model. This dynamic model constituted by the axis of the body-arm cylinders allows the decoding of orders, using either one or two arms, from the analysis of image sequences following the procedure shown in the flow diagram of figure 11.

The elemental gesticular orders initially interpreted by the developed experimental system are: *Advance, Stop, Down, Up, Go-Down, Raise-Up, Increase, Slow* and *Approach*. After the detection of the head and one or two arms, according to their relative position, the system verifies if the movement implies a modification of the arm height with respect to the floor or a change of its distance to the body. Thus, when the two arms are used, a periodic vertical movement changing the *z* value is considered a *Slow* order. On the contrary, if the height is slightly constant at the shoulders and elbow level, and the hand periodically moves horizontally, the command is interpreted as *Advance*.

Commands given by movements maintained for periods longer than one second and with amplitudes over 40 cm are interpreted with notable reliability. Thus the system can become a computer peripheral for the interpretation of elemental commands, those described above, or any other predefined command that can be added.

Hardware configuration

The system has been designed to achieve enough computing speed to make the tracking of most common gesticular commands possible. The developed system consists of two PCs (Fig. 12).

The first computer calculates the operator's position and posture, the one that best fits the images acquired by the different cameras. This computer contains specific image processing hardware that extracts the operator's silhouette from his movement detection in real time and validates its shape by comparing it with the model. This specific processor avoids the need for a high performance workstation.

The estimated movement is supplied to the second computer using its own local area network, this PC is devoted to gesture interpretation. It contains the power interfaces to directly or indirectly control the teleoperated machines and serves as the operator input for system set-up. This computer also calibrates the cameras, when some environment conditions change, and defines the size and tolerance parameters of the generic models

Specific pre-processing unit

A specific computer board for movement detection and segmentation, which operates over the PC bus, was designed to avoid the need to use expensive computer equipment. The hardware processor was specifically designed to extract the pixels in the image that correspond to moving points in the setting. Its function is the implementation of the gradient vector comparison in real time. The operating time is 20 or 40 miliseconds depending on the images used, a single frame or a complete image.

The processor consists of two main modules, based on a pipe-line structure, which operate at pixel level. The first one implements movement segmentation from gradient comparison, and the second carries out movement detection by

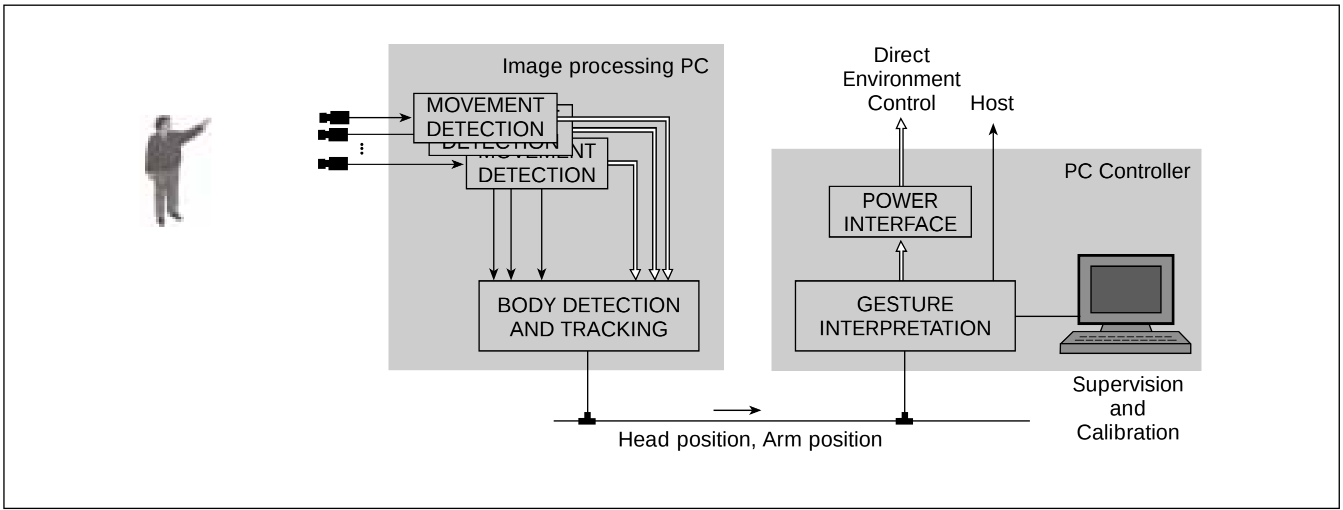


Figure 12. System architecture

adapting this segmentation to movement dynamics. Fig. 13 shows a diagram of the processor structure.

The input image I_i is compared with the previous reference image I_{i-k} following two parallel paths. The first implements image movement segmentation by means of comparing the gradient vectors G_t and G_{t-k}. Initially, since the image is acquired serially, pixel by pixel, it is necessary to carry out a serial to parallel conversion to feed the digital signal processor (DSP) that implements the gradient comparison. Then, the segmentation results are stored in an alternate way in one of the two memories accessible to the bus. The use of alternative memories allows the computer to read the segmentation results at the same time as another image is being segmented, without the need to use a FIFO memory.

The second path used implements movement detection. First, the brightness difference between homologue points in two successive images is computed and compared with a threshold. The next module rejects the windows, or image areas, containing a number of segmented pixels of less than a predefined threshold. These isolated pixels are considered to be due to noise. When there are windows containing a significant number of moving pixels, a signal (PC Interrupt) is sent to the computer to request the analysis of the segmented image.

The movement detection circuitry, consisting basically of logical functions, is implemented by means of standard programmable devices. The controller logic that generates the functions necessary to decode the control addresses, to maintain the state control registers, to control the memory input/output logic as well as the timing, synchronization and

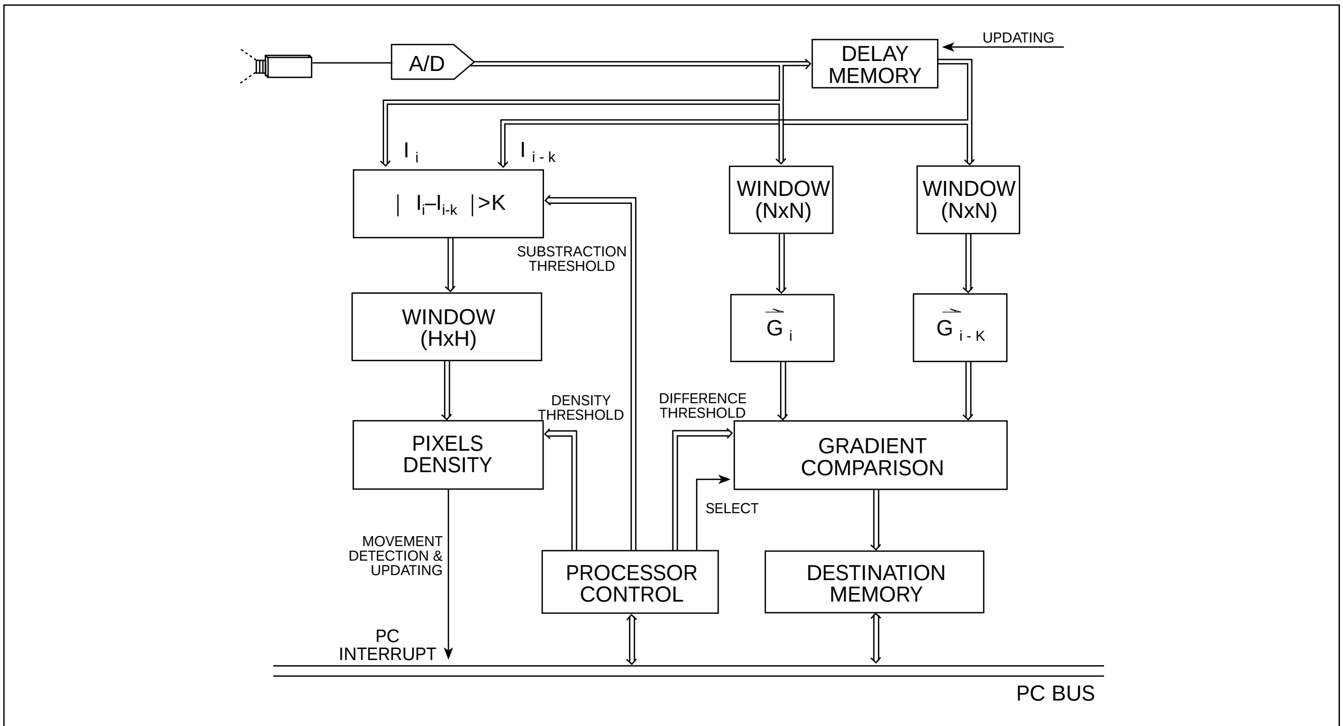


Figure 13. Specific processor

the access to the bus protocol is implemented in the same way using FPGA devices.

Camera configuration and calibration

The number of cameras placed around the setting is variable according to the application and the environment. Their configuration is essential to obtain an adequate setting perspective to avoid occlusions, to have the adequate distance to the human operator and to acquire images free from disturbances produced by other objects.

Although only two stereo-cameras are necessary to detect and track the moving objects in a tridimensional space, the use of multiple cameras (Fig. 14) allows the selection of two which, at every instant *t*, provide the best data to identify the significant body points that define human posture. This selection is done automatically by analyzing the images and choosing the pair that provides the best 3D measurements. A supervisor module continuously evaluates the convenience of changing from one camera to another, from the distance evaluation, the error and the visual area without occlusion parameters, using a hysteresis level to minimize camera changes. The presence of moving objects, other than the expected operator, produces more features to be analyzed and classified, thus introducing more misclassification risks and requiring longer computing time. The operator can be partially or totally occluded from the field of vision of some of the cameras depending on his position relative to other elements in the setting. He can even occlude part of his body when giving a gesticulative command. Finally, the distance to the camera is important to reliably detect the singular points of the image and to fit the model to the estimated body posture. A deficient distance can hinder the visualisation of some relevant body parts, while operating too far from the cameras produces very low resolution for feature extraction and classification. The adequate selection of the operative cameras can reduce such situations.

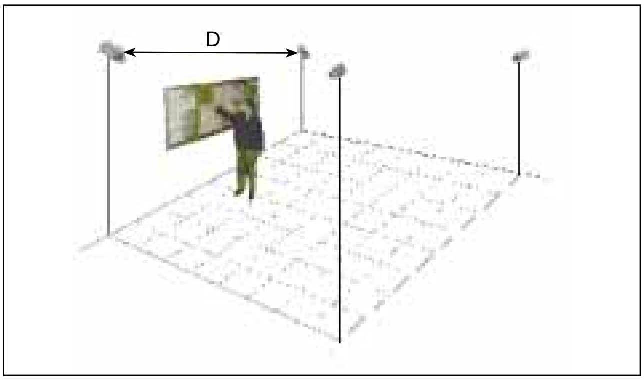


Figure 14. Camera configuration

Camera calibration is essential for 3D measurement. Calibration techniques usually consist of previously identifying a set of parameters: intrinsic, those that depend uniquely on the characteristics of the camera itself such as distortion coefficient, scale factors and the optical center; and extrinsic, those that correspond to the camera position and orientation [13, 48].

In the described system the intrinsic parameters are previously calibrated in the laboratory using a reticular pattern of a known size. From this calibration, a non-linear function $\Phi(X_I,Y_I)$ is determined which, together with the coordinates of a pixel of the image, provides the three components of a director vector (V_C). This vector defines the beam passing through the focus in the direction marked by the setting point, corresponding to the pixel, in space coordinates (Fig. 15).

The inverse function $\Phi^{-1}(X_c,Y_c,Z_c)$ is likewise computed to provide the position of a point in the coordinates space relative to the reference camera system *{C}*. This function provides the pixel coordinates in the image plane.

The homogeneous transformation matrix ${}^{C}T_W$ converts a point *P* in the world or setting coordinates frame *{W}* to the relative camera frame, while the inverse matrix ${}^{W}T_C$ implements the opposite function. Consequently, from the matrix

$$ {}^{w}T_c = \begin{pmatrix} r_1 & r_2 & r_3 & T_x \\ r_4 & r_5 & r_6 & T_y \\ r_7 & r_8 & r_9 & T_z \\ 0 & 0 & 0 & 1 \end{pmatrix} $$

a point can be referred to another frame:

$$ \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}_w = {}^{w}T_c \cdot \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}_c $$

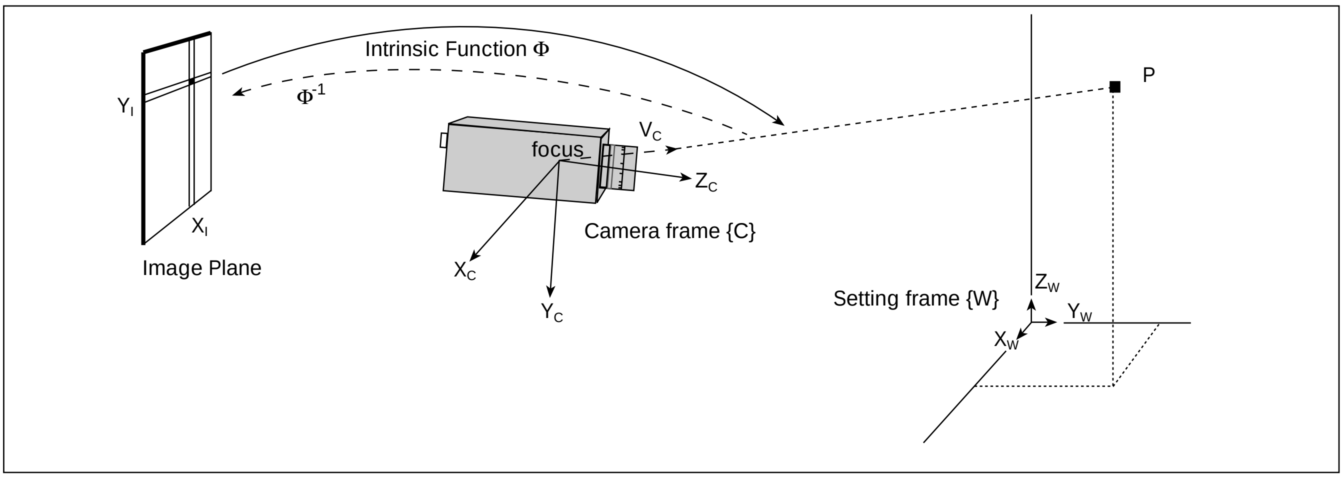


Figure 15. Camera calibration frames

Where vector (T_X, T_Y, T_Z) represents the camera position with respect to the setting frame and the values r_1 to r_9 constitute the components of a rotation matrix ${}^{W}R_C$ (3x3) that defines the camera orientation. This rotation matrix is completely defined by means of the three rotation angles (roll, pitch and yaw), the Euler angles, or using another reference base.

The calculation *in situ* of the six extrinsic parameters for each camera requires the use of reference calibration points. Some assumptions are taken to simplify the calibration process and to reduce the number of reference points. These assumptions are based on the fact that the effects of possible position errors in stereoscopic systems with a high baseline (of about several meters) are much less than those produced by orientation errors. Thus, rejecting the camera positioning errors, only two points, non-colinear with the camera center point and far enough from each other, are necessary to estimate the three camera orientation parameters with sufficient precision.

The effect of an error in the camera positioning produces a triangulation position error that can be considered constant and consequently can be adjusted as an offset error. Provided that the reference calibration points, given manually, are precise, the relative position of the cameras supplied by the user is less critical. Assuming that the camera position is measured with an error less than 2 cm, the system iteratively recalculates this position from the reference calibration points.

By increasing the number of calibration points it is possible to increase triangulation precision, by averaging the errors due to the calibration point measurements.

The calibration points can be natural elements in the operating setting, such as windows or door corners. In this case it is not necessary to modify the working environment to introduce artificial landmarks. Nevertheless, the user is required to indicate the position of each selected pixel with respect to setting frame.

Body localization

As in any physical system, there are multiple factors that introduce imprecisions in pixel location in stereovision systems. The analysis of potential errors in target location includes position indetermination and location dynamic errors.

The location error of a detected contour segment is due both to noise and to the digitizing process. Noise produces an error that is usually modeled as a gaussian function and produces some indetermination on the estimation of the pixel position with respect to the real target position. It is demonstrated that the probability that this error together with the digitized one is bigger than ± 2 pixels is almost null [3]. In stereoscopic vision systems, this error produces an indetermination in the position of a point in space less than:

$$\varepsilon_P = \pm d_{max} \cdot tg(2 \cdot \Delta_{pixel}) / \sin^2(\alpha/2 - 2 \cdot \Delta_{pixel}).$$

Where α is the angle formed by the two projection lines of a point to the stereo image planes ($2 \cdot \Delta_{pixel} < \alpha < 90\Upsilon$); d_{max} is the maximum distance from the camera to the object; and Δ_{pixel} is the angle defined by a pixel width measured in the lateral extreme of the image. In this area of the image optical deformation is usually highest.

The definition of an extreme body point from its projection to the image plane depends on the cameras field of vision. Fig. 16 shows the error volume of the detection and triangulation of two singular points, in this case associated to the head. The pixels on the two stereo images corresponding to the most prominent point, are the projection of points P1 and P2. The possible projection points of the head depend on the cameras field of vision and define approximately a half

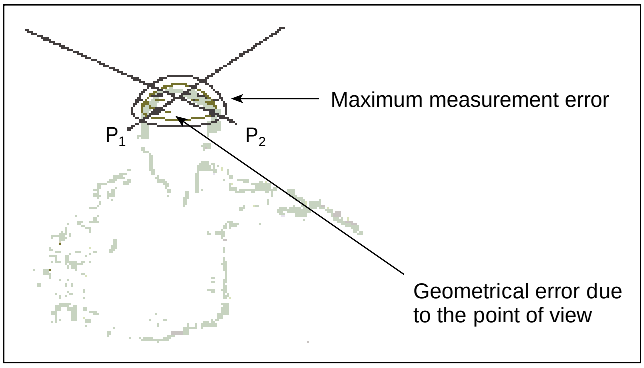


Figure 16. Measurement error of extreme point position

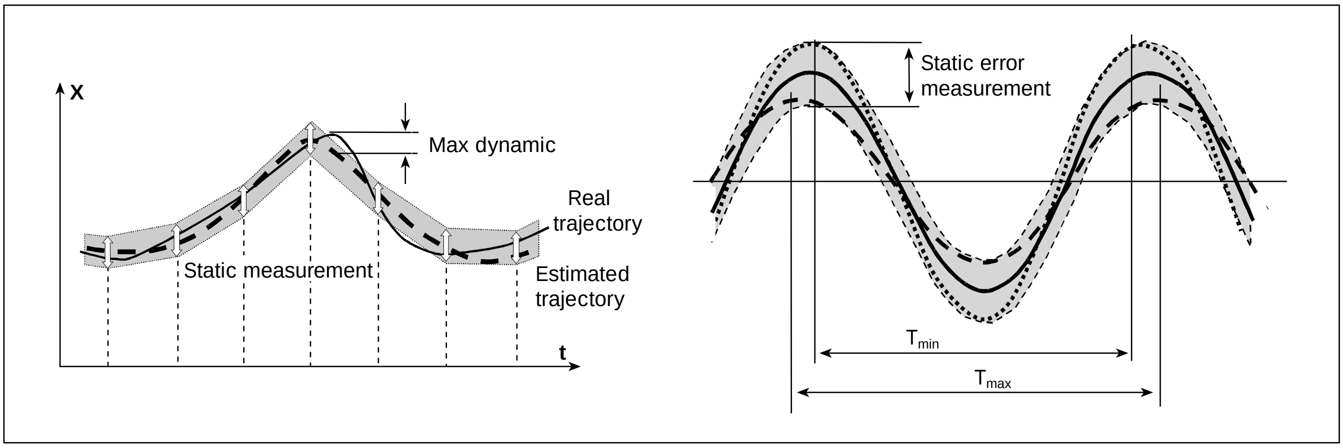


Figure 17. Spatial and frequential errors. a) Dynamic and b) frequency

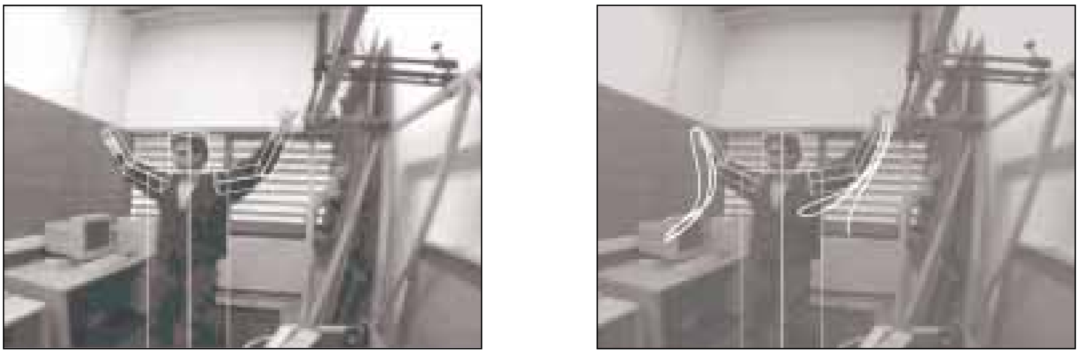


Figure 18. Result of applying the multicylindrical model to a sequence of images

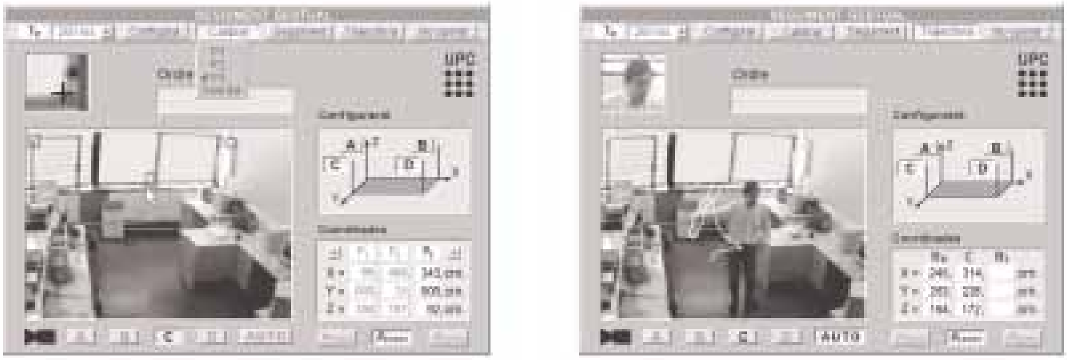


Figure 19. Graphical user interface. a) Calibration mode, b) Tracking mode

ellipsoid, as seen in the figure. The digitizing error enlarges this volume to a second half ellipsoid. In this case, the maximum positioning error is obtained when the direction defined by the pixel with highest curvature and the main ellipsoid axis are parallel. Under this condition, the maximum radium from the two minor ellipsoid radii is the maximum distance boundary. The worst situation occurs when the two cameras detect as extreme point the one located at the maximum distance of the ellipsoids main axis.

If R_Z is the size of main axis, R_X and R_Y are the minor radii of this approximate ellipsoid, then the maximum error ε_S due to the error introduced by the detection of the extreme points is:

$$\varepsilon_S = \sqrt{\max\left(R_x, R_y\right)^2 + R_z^2} + \varepsilon_P$$

this error is not constant and depends on the object to be localized.

Further to static analysis other dynamic errors must be considered. Dynamic location errors (ε_D) are due to the fact that human movement is continuous by nature while it is estimated from discrete samples taken at given time intervals. The minimum time between two consecutive samples, cycle time, T_C, is conditioned by image frequency and the image processing time. However, since the target tracking is carried out from movement detection, this sampling time is increased to fit the operator's dynamics. Thus, T_C is continuously variable, which allows the reduction of the measurement error. The effects of sampling a moving object produces an error in the trajectory estimation as follows:

$$\varepsilon_D(t) = \left| \varepsilon_S + \frac{2\varepsilon_S}{T_c^2} \cdot t^2 \right|$$

and an error in period estimation (ε_T) of an oscillatory movement of amplitude A and frequency f, that can be quantified with the following expression:

$$\varepsilon_F = \frac{2f}{\pi} \arcsin\left(\frac{\varepsilon_S}{A}\right)$$

Fig. 17 shows such errors. In Fig. 17 a) the gray band corresponds to the area where the real target trajectory is expected to pass, and the dashed line is the estimated trajectory from the sampled points and the evaluated static positioning errors. The continuous line shows a potential real trajectory that provides the indicated sampled points, and from both curves the maximum error is detected. In Fig. 17 b), which shows the expected band around the estimated trajectory, two potential real trajectories with extreme frequencies, dashed lines, show how different body movement frequencies can produce the same estimated trajectory.

Applying this system to an operation room of 4 x 8 meters, these errors the obtention of the 3D singular point position with a precision over the 2 cm, when a finger is pointing at a target in the center of the image. When the operator is pointing at a distant target, the arm direction, computed as an extrapolation of the elbow-finger line, is obtained within an uncertainty cone with an angle up to 10º. This limits the capability of signaling small distant objects.

The system thus, is more efficient if used to interpret commands given directly with the hand close to the objects within the setting or commands through gestures.

Results and conclusions

From the results obtained with the described experimental system it can be demonstrated that it is possible to reliably interpret a given number of gesture orders and consequently the system offers a wide field of applications. Fig. 18 shows a sequence of an operator's movement and the tracking of his body and hand trajectory.

The interpreted gestures are common enough so as to be possible for different human operators to give the required commands in a very intuitive way, and consequently without requiring a specific learning process. In this way, it is possible to include human gestures as an additional means of MMI which is suitable for in industrial environments, in applications in which the access to electrical or conventional computing supports can incur a waste of time or difficulties derived from the environment. The possibilities to incorporate gestures to provide complementary information to the oral communication are also very promising. For instance, to point at an object while orally describing its characteristics. In this respect, the work developed at the MIT Artificial Intelligence Laboratory within the project *Intelligent Room* [14] are well-known. A first version of the described system was incorporated into this laboratory, to be used as a digital pointing device based on operator gestures. To operate in a robust enough way using this gesture interface, an autocalibration system that enables system operation with minimal user intervention is necessary. Therefore, a MMI graphical interface was designed. The interface allows the user to define some windows over the image in which some reliable reference points can be extracted to continuously recalibrate the cameras. In this case, the user must enter the X, Y, Z coordinates of the selected points, the corners of the windows in this example, Fig. 19 (a), in the phase previous to calibration. After this operation, the obtained measures, Fig. 19 (b) are absolute values which are continuously recalibrated to compensate the possible movements of the camera positions.

Since the system tracks the operator when he moves within the setting and adjusts the obtained image to a simplified polycilindrical geometrical model, it is possible to automatically switch to the two best placed cameras to obtain the two stereo images that allows the computation of the best 3D information about the operator at each instant. From these images it is also possible to obtain a stabilized image of the zone in which the operator's face appears thus providing this data as additional information.

These studies and the obtained results offer wide possibilities of human-machine interaction, in a natural way, without needing specific equipment such as gloves, sticks or other effectors or instruments. Such interaction can be used not only in industrial and professional environments but also in assistance applications for people with perception limitations or for elderly people who have a psychological dislike of conventional computer aids.

References

[1] O.A. Alsayegh and D.P. Brazakovic (1998). Guidance of Video Data Acquisition by Myoelectric Signals for Smart Human-Robot Interfaces. In Proc. Int. Conf. on Robotics and Automation.

[2] J. Amat, A. Casals, M. Frigola (1999). Stereoscopic System for Human Body Tracking in Natural Scenes. ICCV Workshop on Modelling People, MPEOPLE'99.

[3] J. Aranda (1997). Aportació als mètodes de seguiment tridimensional d'objectes d'alta velocitat d'operació, mitjançant l'estereovisió. Tesi Doctoral Universitat Politècnica de Catalunya, Barcelona.

[4] A.Azarbayejani, C. Wren and A. Pentland (1996). Real-Time 3-D Tracking of the Human Body. Media Lab. Technical Report 374, MIT, Massachusetts.

[5] A. Azarbayejani and A. Pentland (1996). Real-time self-calibrating stereo person tracking using 3-D shape estimation from blobs features. In Proceedings IEEE Conference on Computer Vision and Pattern Recognition, Vienna.

[6] R. Azuma, G. Bishop (1994). Improved Static and Dynamic Registration in an Optical See-through HMD. In Proc. SIGGRAPH, Orlando.

[7] E.D. Bernardo, L. Goncalves and P.Perona (1998). Monocular Tracking of the Human Arm in 3D. Computer Vision for Human-Machine Interaction. Ed. by R. Cipolla and A. Pentland. Cambridge University Press.

[8] M. Bichsel (1991). Strategies of Robust Object Recognition for the Automatic Identification of Human Faces. PhD thesis, ETH, Zurich.

[9] M.J. Black and Y. Yacoob (1995). Tracking and Recognizing Rigid and Non-rigid Facial Motions using Local Parametric Model of Image Motion. Proc. ICCV.

[10] C. Bregler and J. Malik (1997). Video Motion Capture. Technical Report CMU-CS-97-973, University of California, Berkeley.

[11] R. Brunelli and T. Poggio (1993). Face recognition: Features versus templates. IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(10): 1042-1052.

[12] P.J. Burt (1989). Multiresolution techniques for image representation, analysis, and 'smart' transmission. In

SPIE vol.1199, Visual Communications and Image Processing IV, pp. 2-15.
[13] F. Chaumette (1990). The relation vision-command: Theory and applications to robotic tasks. Ph.D Thesis Rennes, France.
[14] M.H. Cohen (1998) Design Principles for Intelligent Environments. Proc. AAAI-98.
[15] G.W. Cottrell and J. Metcalfe (1991). EMPATH: Face, Gender and Emotion Recognition Using Holons. Advances in Neural Information Processing Systems, vol.3, pp. 564-571.
[16] T. Darrel and A. Pentland (1993). Space-time gestures. In Proc. IEEE Conf. on CVPR.
[17] J.W. Davis and A.F. Bobick (1997). The representation and Recognition of Action Using Temporal Templates. IEEE Conference on Computer Vision and Pattern Recognition.
[18] M. Dhome, A. Yassine and J. Lavest (1993). Pose estimation of a robot arm from one grey level image. Workshop on Computer Vision for Space Applications, pp. 298-310.
[19] I. Douros, L. Dekker, B.F. Buxton (1999). An Improved Algorithm for Reconstruction of the Surface of Human Body from 3D Scanner Data Using Local B-Spline Patches. Workshop on Modelling People, MPEOPLE'99.
[20] D.M. Gavrila and L.S. Davis (1996). 3-D model-based tracking of humans in action: a multi-view approach. IEEE Computer Vision and Pattern Recognition.
[21] P.W. Hallinan (1994). A low-dimensional representation of human faces for arbitrary lighting conditions. In Proceedings IEEE Conference on Computer Vision and Pattern Recognition, pp. 995-999.
[22] I. Haritaoglu, D. Harwood and L.Davis (1998). Ghost: A human body part labelling system using silhouettes. International Conference on Pattern Recognition, 16(20), Brisbane.
[23] D. Hogg (1983). A program to see a walking person. Image Vision Computing, 5(20).
[24] B.K.P. Horn and B.G. Schunk (1981). Determining optical flow. Artificial Intelligence, 17:185-203.
[25] C.L. Huang and C.W. Chen (1992). Human facial feature extraction for face interpretation and recognition. Pattern Recognition, 25(12), pp. 1435-1444.
[26] K. Ishibuchi, H. Takemura and F. Kishino (1992). Real Time Hand Shape Recognition for Man-Machine Interfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1408 -1412.
[27] S. Iwasawa, J. Ohya, K. Takahashi, T Sakaguchi, S. Kawato, K. Ebihara and S. Morishima (1999). Real-time 3D Estimation of Human Body Postures from Trinocular Images. ICCV Work. Modelling People, MPEOPLE'99.
[28] I.A. Kakadiaris, D. Metaxas (1995). Human Body Model Acquisition from Multiple Views. IEEE Computer Vision and Pattern Recognition Conference, pp. 618-623.
[29] M. Livingston and A. State (1995). Improved Registration for Augmented Reality Systems via Magnetic Tracker Calibration. Tech. Report TR95-037 of University of North Carolina.
[30] Y. Luo, F.J. Perales and J.J. Villanueva (1992). An Automatic Rotoscopy System for Human Motion Based on a Biomechanical Graphical Model. Computer and Graphics, Vol. 16 (4), pp. 355-362.
[31] C. Maggioni and B. Kämmerer (1998). Gesture Computer - History, Design and Applications. Computer Vision for Human-Machine Interaction. Ed. by R. Cipolla and A. Pentland. Cambridge University Press.
[32] D. Metaxas and D. Terzopoulos (1993). Shape and nonrigid motion estimation through physics-based synthesis. IEEE Trans. on Pattern Analysis and Machine Intelligence, 15(6).
[33] H. Nugroho, J. Hwang and S. Ozawa (1994). Tracking Human Motion in a Complex Scene Using Textural Analysis. IECON 94, pp. 727-732.
[34] H. Ohta, H. Saji and H. Nakatani (1998). Recognition of Facial Expressions Using Muscle-Based Feature Models. Proc. ICPR vol. 2.
[35] J. Ohya, T. Miyasato, R. Nakatsu (1999). Virtual Reality Technologies for Multimedia Communicatios. In Mixed Reality - Merging Real and Virtual Worlds. Ed. Y. Ohta, H. Takamura. Springer-Verlag, NY.
[36] E. Ong and S. Gong (1999). Tracking Hybrid 2D-3D Human Models from Multiple Views. ICCV Workshop on Modelling People, MPEOPLE'99.
[37] A. Pentland and B. Horowitz (1991). Recovery of nonrigid motion and structure. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(7): 730-742.
[38] A. Pentland, B. Moghaddam and T. Starner (1994). View-based and modular eigenspaces for face recognition. In Proceedings IEEE Conference on Computer Vision and Pattern Recognition, pp 84-91.
[39] F.J. Perales and J.Torres (1994). A System for Human Motion Matching between Synthetic and Real Images Based on a Biomechanical Graphical Model. IEEE Workshop on Motion of Non-Rigid and Articulated Objects.
[40] R. Polana and R.Nelson (1994). Low level recognition of human motion. In Proc. of IEEE Workshop on Motion of Non-Rigid and Articulated Objectes, pp. 77-82.
[41] R. Rashid (1979). LIGHTS: A study in motion. In Proc. DARPA Image Understanding Workshop, pp. 57-68, Nov, 1979.
[42] P.Remagnino, A. Baumberg, T. Grove, D. Hogg, T.Tan, A.Worrall and K.Baker (1997). An Integrated Traffic and Pedestrian Model-based Vision System. In Proc. of BMVC97, vol. 2, Colchester.
[43] J.O'Rourke and N.I. Badler (1980). Model-based image analysis of human motion using constraint propagation. IEEE Trans. Pattern Analysis and Machine Intelligence, 2(6), pp. 522-536.
[44] M.A. Shackleton and W.J. Welsh (1991). Classification

of facial features for recognition. In Proceedings IEEE Conf. on Computer Vision and Pattern Recognition.

[45] T. Starner and A. Pentland (1995). Visual Recognition of American Sign Language Using Hidden Markov Models. In Proc. of Int. Workshop on Automatic Face and Gesture Recognition, Zurich.

[46] D. Terzopoulos and K. Waters (1993). Analysis and Synthesis of Facial Images Sequences Using Physical and Anatomical Models. IEEE Trans. on PAMI 15(6).

[47] D. Thalmann, J. Shen and E. Chauvineau (1996). Fast Realistic Human Body Deformations for Animation and VR Applications. In Computer Graphics International.

[48] R.Y. Tsai (1987). Aversatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-self TV cameras and lenses. Journal of Robotics and Automation, vol. 4, pp. 323-331.

[49] M. Ward, R. Azuma, R. Bennett, S. Gottscalk, H. Fuchs (1992). A Demonstrated Optical Tracker with Scalable Work Area for Head-Mounted Display Systems. In Proc. of the Symposium on Interactive 3D Graphics, pp. 43-52.

[50] J.J. Weng and Y.Cui (1998). Recognition of Hand Signs from Complex Backgrounds. Computer Vision for Human-Machine Interaction. Ed. by R. Cipolla and A. Pentland. Cambridge University Press, pp. 235-265.

[51] C. Wren, A. Azarbayejani, T. Darrel and A. Pentland (1995). Pfinder: Real-time tracking of the human body. In Pothonics East, SPIE Proceedings Vol. 2615, Bellinham.

[52] A. Wright, M. Stanisic (1990). Kinematics Mapping between the EXOS Handmaster Exoeskeleton and the Utah/MIT Dextrous Hand. IEEE Int. Conf. on System Engineering, Pittsburgh.

[53] H. Wu, T. Shioyama and H. Kobayashi (1998). Spotting Recognition of Head Gestures from Color Images Series. Proc. ICPR vol. 1.

[54] M. Yachida and Y. Iwai (1998). Looking at Human Gestures. Computer Vision for Human-Machine Interaction. Ed. by R. Cipolla and A. Pentland. Cambridge University Press.

[55] M. Yamamoto and K. Koshikawa (1991). Human motion analysis based on a robot arm model. IEEE Conf. Coputer Vision and Pattern Recognition, pp. 664-675.

[56] A.L. Yuille, P.W. Hallinan and D.S. Cohen (1992). Feature extraction from faces using deformable templates. International Journal of Computer Vision, Vol. 8(2), pp. 99-111.

[57] J.Y. Zheng and S. Suezaki (1998). A model Based Approach in Extracting and Generating Human Motion. In Int. Conference in Pattern Recognition.

About the authors

Josep Amat (Barcelona 1940). Dr. Amat graduated in Electrical Engineering in 1967 from the School of Engineering of Barcelona, and obtained his Ph. D. in 1977 from the Universitat Politècnica de Catalunya (UPC). In 1968 he joined UPC to teach Digital Systems. He is currently teaching several subjects in the field of computer technology and robotics at the Automatic Control and Computer Engineering Departament (UPC). His research started in signal processing, and this work derived to computer vision in 1980, and was mainly oriented to its application to inspection and advanced robotics systems. He is a leading research in computer vision at UPC. In 1995 he joined the Institute of Robotics (UPC/CSIC). His main research projects deal with underwater and medical robotics. Man-machine interaction in these applications implies the design of friendly interfaces constituting another research area related to the work presented. Member of the Institut d'Estudis Catalans.

Alícia Casals (Barcelona 1955). Dr. Casals graduated in Electrical Engineering in 1977 from the School of Engineering of Barcelona (ETSEIB-UPC) and completed her Ph. D. in Computer Science in 1983 (FIB-UPC). She is now a lecturer in the Automatic Control and Computer Engineering Department at the (UPC). Her research is devoted to the field of Robotics and Computer Vision, mainly oriented to robotics applications, since is1984 responsible for the research line in robotics at UPC. She is currently leading projects on robotics applied to the medical field, both in rehabilitation and in surgery. In both fields the interfaces with non-computer specialists have led to the study of new and more friendly and effective communication methods, the study on gesture interfaces is a response to this need.

Manel Frigola (Blanes 1966) received his M.S.E. degree in Computer Engineering from the Computer Science Faculty in 1992, from UPC. The study presented is part of his doctoral research, done under the direction of Dr. Amat. He is now an Assistant Lecturer of Control Engineering at UPC. His current research interests include computer vision, robotics and man-machine interaction.

Jaume Pagès (Girona 1946). Background in Teaching, Industrial Engineering and Computing. Ph.D in Industrial Engineering in 1975. He is now a professor at UPC, where he has been lecturing in the fields of Mechanics, Mathematics and Automatic Control since 1971. His research is in the areas of signal processing, linear and non-linear filter synthesis, analysis and solving of control problems and planning of interplanetary missions. Vice-rector at UPC for Academic Organization from 1986 to 1994. Rector from 1994, re-elected in 1998. Member of the Institut d'Estudis Catalans.