

# Animating H-Anim using real image sequences

J.M. Buades, F.J. Perales  
Computer Graphics and Vision Group  
Department of Computer Science  
Universitat de les Illes Balears (UIB)  
e-mail: vdmijbr3@ps.uib.es, paco@anim.uib.es

## Abstract

*The study of the movement of the human body is applicable to many current fields of science and technology, such as biomechanical study in sporting areas, the study of functional disabilities, the integration of people in virtual worlds, human interaction with computers and the ability to recognise people in scenes.*

*This article presents a system to analyse and synthesise human movement by means of computer vision techniques using a model of a person based on the standard humanoid H-Anim. The main aim of the system is to carry out a correspondence between this graphic model and the person in movement in real images. The process is intended to be non-invasive and automatic although it is currently used with the manual intervention of the user. The final result will enable us to integrate the synthetic model and its movement with the real person in a real or virtual world adaptable to different applications.*

## 1.- Introduction

The analysis of the movement of the human body may be approached from different perspectives depending on the type of application to be considered. In our case, the techniques used are conditioned according to the initial hypotheses of minimum perturbation of movement and/or surroundings, using exclusively visual information of the scene and a biomechanical model of the person. With this aim in mind, in order to obtain the parameters of human movement we use computer vision techniques (pre-processing, segmentation, matching of entities, 3D reconstruction etc.), capturing the individual from an arbitrary number of cameras which enable us to obtain as much information as possible. This part will be explained in detail in the following section. In the third section, from the information obtained from the cameras, we determine the position of the person's articulations; this process is called matching. By means of the positions of the articulations, we either place the humanoide H-Anim over the image captured or we substitute the image captured for the virtual world modelled on VRML. In the fourth section we explain the 3D representation used, while in the fifth section we present the results together with the conclusions and future work.

## 2.- Capturing Process

In the capturing process we have two possibilities; to use two synchronised colour cameras, or four synchronised black and white ones. There is no synchronisation between the colour ones and the black and white ones, although through the network it is possible to synchronise with a maximum of 33 milliseconds. If we want to use the six cameras and

thus obtain a more exact reconstruction of the movement, this can be done by restricting the movement of the person to non-rapid movements. Figure 1 shows the capturing system; the synchronisation between the two systems is carried out via an Ethernet.

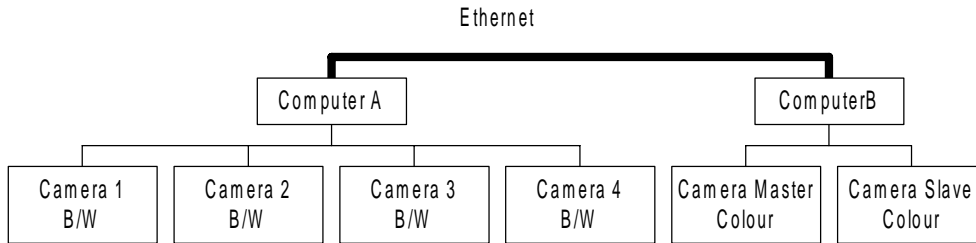


Figure 1. Configuration of cameras

However, by using a group of cameras the aperture of the lens is carried out at the same instant in time  $t$ , thus keeping all the cameras synchronised. The capture is carried out at 30 images per second and at a resolution of 640x480. If one wants to record the movement for posterior analysis it is necessary to keep it in memory and afterwards transfer it into a folder. This limits the recording time to 10 seconds, around 384 MBytes of RAM memory available. If the processing is carried out in real time there is no time limit.

As far as the calibration of the cameras is concerned, we used a basic algorithm which takes into account all the usual intrinsic and extrinsic parameters although initially we are not dealing with any kind of distortion, since the graphic producing packs such as OpenGL do not take into account the distortions that are unable to be represented through a matrix transformation. Either way, the optics used do not introduce any appreciable distortion.

### 3.- Matching

The matching process associates each articulation with a 2D point in the image. This process consists of analysing each image obtained from the cameras in an instant of time  $t$ . Once we have the articulation located in two or more cameras we estimate the most accurate 3D point; an articulation may only be detected in one or no image so the process will have to be completed with contextual information from a much higher level.

What is intended is to obtain the most accurate approximation possible, thus the virtual human must have similar anthropometric sizes to a person, In order that the humanoid H-Anim has the same size, we use our own editor [1], figure 3 shows an adjustment of the humanoid to the real size of the person. For the measurement, the person takes up four classic ergonomic postures which enable us to obtain the maximum information, without taking into account the size of the phalanges, the fingers, or the vertebrae.



Figure 2. Humanoid adjusted to the size of the person

On analysing the sequence, we are able to apply physical and temporal restrictions which help us to carry out the matching whilst reducing the errors and the search space [4]. This adjustment process is conditioned by a set of conditions which are optimised in each case and type of movement. The restrictions are:

- Angle and distance limitations of the articulations
- Temporal continuity of the movement in speed or acceleration
- Prediction of the movement based on a database of known movements, in the case of having non-visible articulations.
- Collision of entities.

This set of restrictions defines a level of matching that can follow two criteria and currently the first of Figure 3 is the one that is used. This matching process currently works in a semiautomatic way, thus making it possible to work at different levels of precision depending on the application for which the results are required.

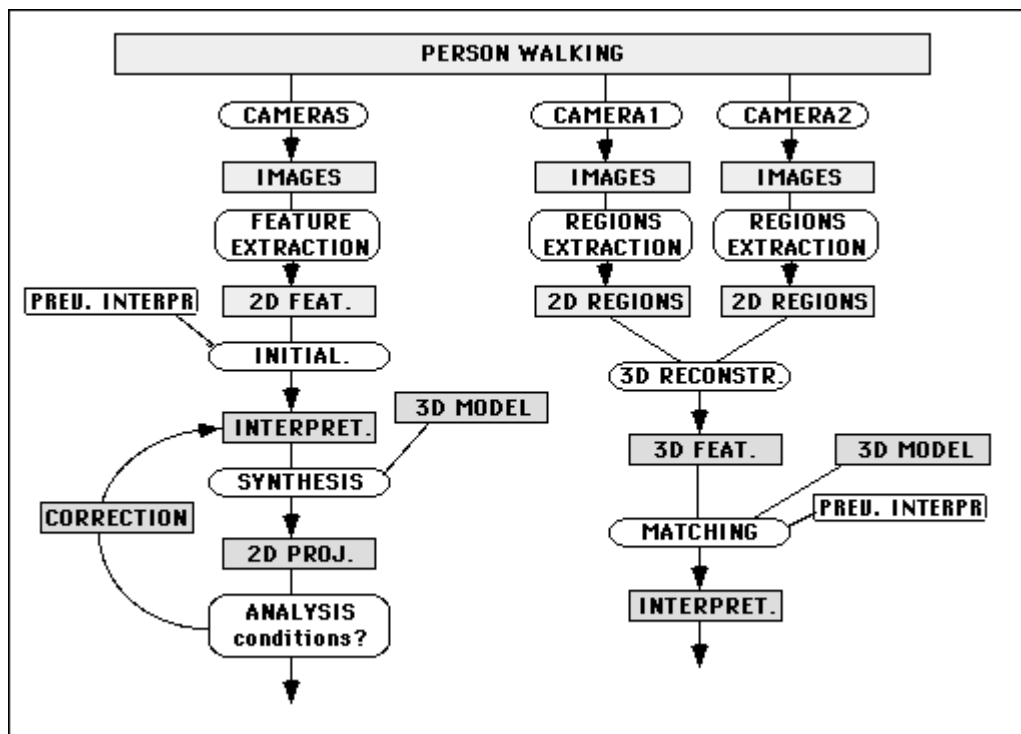


Figure 3. Our proposed system correspond to the left system of the diagram

At the moment the automatic detects the occupied volume, computed from the  $n$  different cameras, this is done in real time and at rate of 30 fps (cameras capture speed, NTSC). To compute it we do the following steps:

1. The study volume is divided in equal volume voxels.
2. For each voxel, compute if it is occupied or not.

This process is carried out for each captured frame, and for the total volume, but normally the volume that the subject occupies is lower than the total volume, therefore we can decrease computing time if we limit the study volume to a reduced volume. For this if in the previous frame any voxel has been detected occupied, the algorithm computes the bounding box that contains the occupied volume, and only computes for this subset volume, the bounding box. We use the following algorithm.

```
Compute the study limits in previous frame
For each voxel in bounding box
    Compute if the voxel is occupied or not
While any of the bounding box border has any occupied voxel, and not is a total
    volume border
    Expand the bounding box one unit in the occupied border
    For each new added voxel
        Compute if the voxel is occupied or not
```

This allows us to restrict the studied volume to the zone of interest and modifying and moving the bounding box according to the movements of the subject.

Until now we have explained how to compute the volume in general, but we have not explained how to determine if a voxel is occupied or not. A voxel is occupied if the voxel has changed in all cameras, for this process we use a background image that will serve as reference, with this reference image and the captured frame we carry out a subtraction and applied and adaptive threshold. We then have to discern two cases, colour cameras and black and white cameras. In the b/w cameras the colour space is one dimensional, the light intensity captured from the camera, therefore we have little useful information. In this case we only can apply a difference.

$$abs(frame(x, y) - background(x, y)) \geq threshold$$

This causes the appearance of non desired shadows if we have a low threshold. And we will not detect interesting parts if the threshold is high. In colour cameras the colour spaces is 3D thus we have more information than we can use. After doing some tests with different colour spaces we have finally chosen the HSI colour space, rather than using YUV, RGB or nRGB [9]. Now we have a typically threshold for each component H, S and I. This colour space allows us to eliminate shadows caused by the subject and retrieve information about zones that in b/w colour space was not possible.

To mark a voxel as changed from a camera we do a independent threshold for each component and in at least one of them there should be a change higher than the threshold.

The threshold has been chosen heuristically. The Hue component is cyclic, so the distance in H between 0 and 359 is 1 degree, for this reason we have to check for the component H..

$$\begin{aligned}abs(frame\_I(x, y) - background\_I(x, y)) &\geq threshold\_I \\abs(frame\_S(x, y) - background\_S(x, y)) &\geq threshold\_S \\abs(frame\_H(x, y) - background\_H(x, y)) &\geq threshold\_H \\abs(frame\_H(x, y) - background\_H(x, y)) &\geq 360 - threshold\_H\end{aligned}$$

To achieve the best results we previously smooth the captured and background image, smoothing gives as a result the stabilisation of the H-component and therefore divide by two the threshold of such component. For the I-component, we can allocate a high threshold and thus eliminate shadows without erasing parts of interest of the subject. Finally, the S-component provides us little useful information. The task of matching now focuses in determine over the segmented images and the generated volume which is each of the parts of interest [5], [6], [7] and [8]. At the moment, no high level information is used in this last presented approach. We are working to combine the process presented in the first part of this section with the last segmentation and voxel occupancy criteria.

#### **4.- 3D representation**

When we edit a humanoid or we carry out interactive manual matching, it is very helpful to have the person we are using as reference and vary the measurements of the humanoid actively, therefore we have placed the captured image as a background.

We create a 3D representation of the humanoid with the captured image placed as background, using OpenInventor (3D edition library that uses OpenGL primitives). As the definition of the objects and their parameters in VRML are very similar to OpenInventor, the conversion from the H-Anim humanoid is very easy. One of the principal objectives is the portability of the representation and the animations.

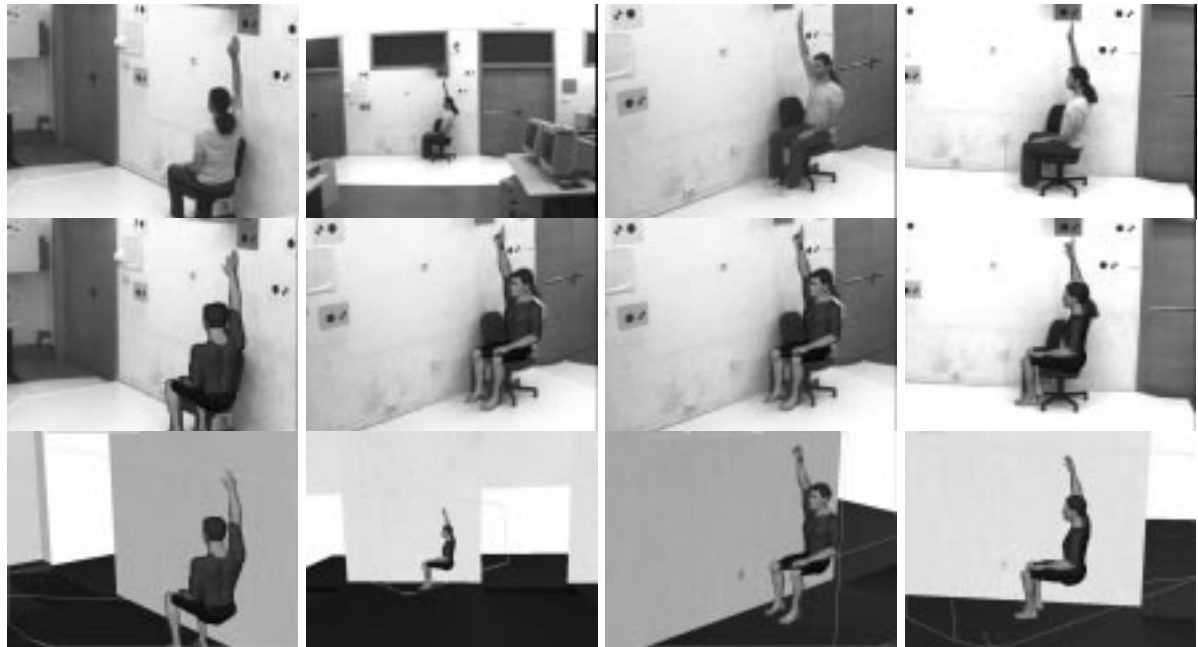
The same way, the scenery or the laboratories are modelled in VRML, which allows us to mix real objects with virtual scenes. This gives us the opportunity to integrate the virtual humanoid in a VRML scenery, such as a sports track in which the humanoid reacts to the movements of the modelled person.

In the automatic process still in progress and is a changeling topic, the representation of the computed volume was done at the beginning by drawing the face of each one of the border voxels between the occupied and non-occupied volume. Now, it is done via Marching Cubes algorithm, which gives us a more accurate representation. We would like to use this information to help to the high matching process as a bounding approximate volume. Of course, try to mach the surface model with a bio-mechanical tree person model is ver complex or may be impossible.

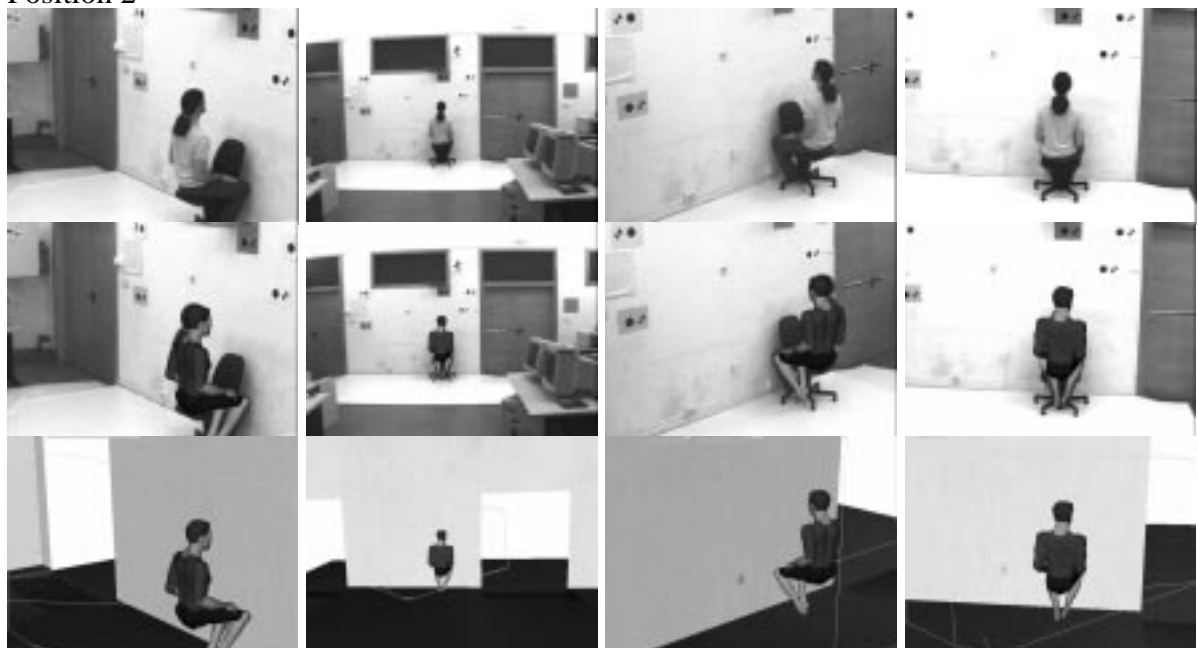
#### **5.- Results**

In this section, we present some examples of sequences colour images. Measurement captures from a person and posterior humanoid adaptation to the subject.

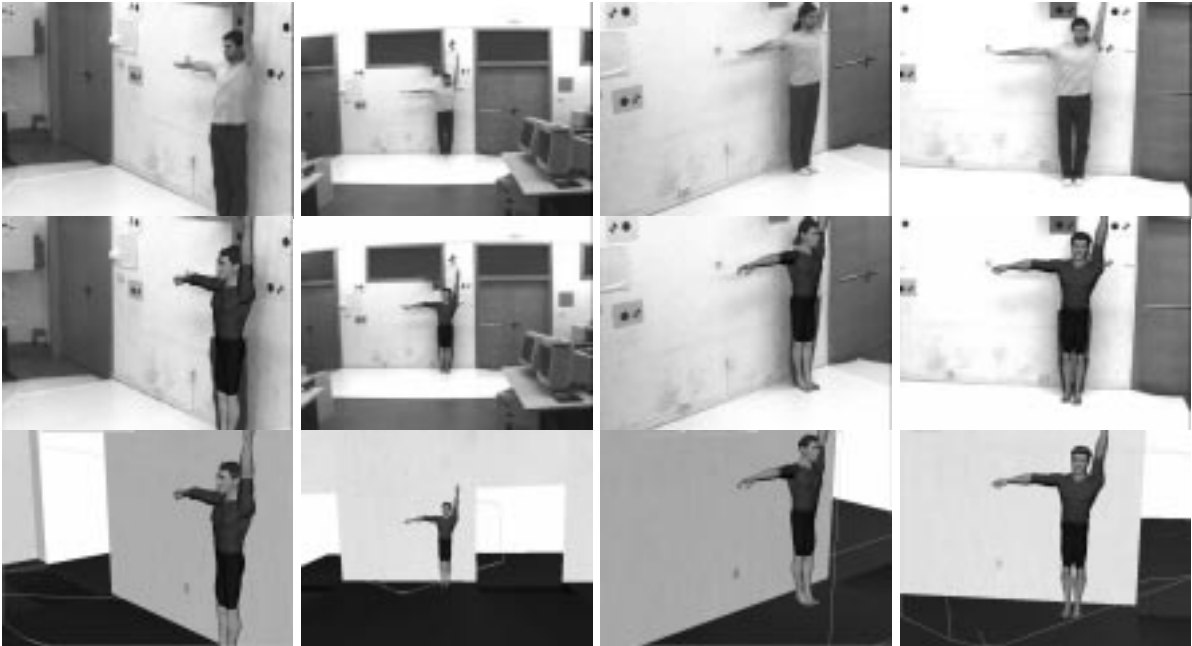
Position 1



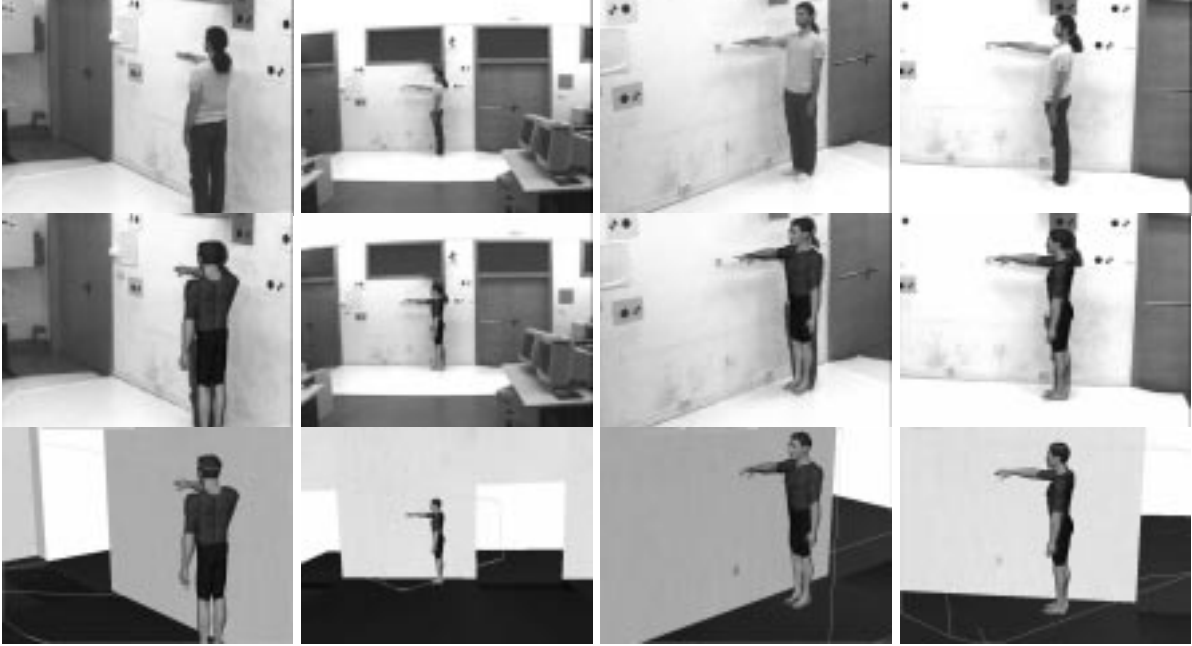
Position 2



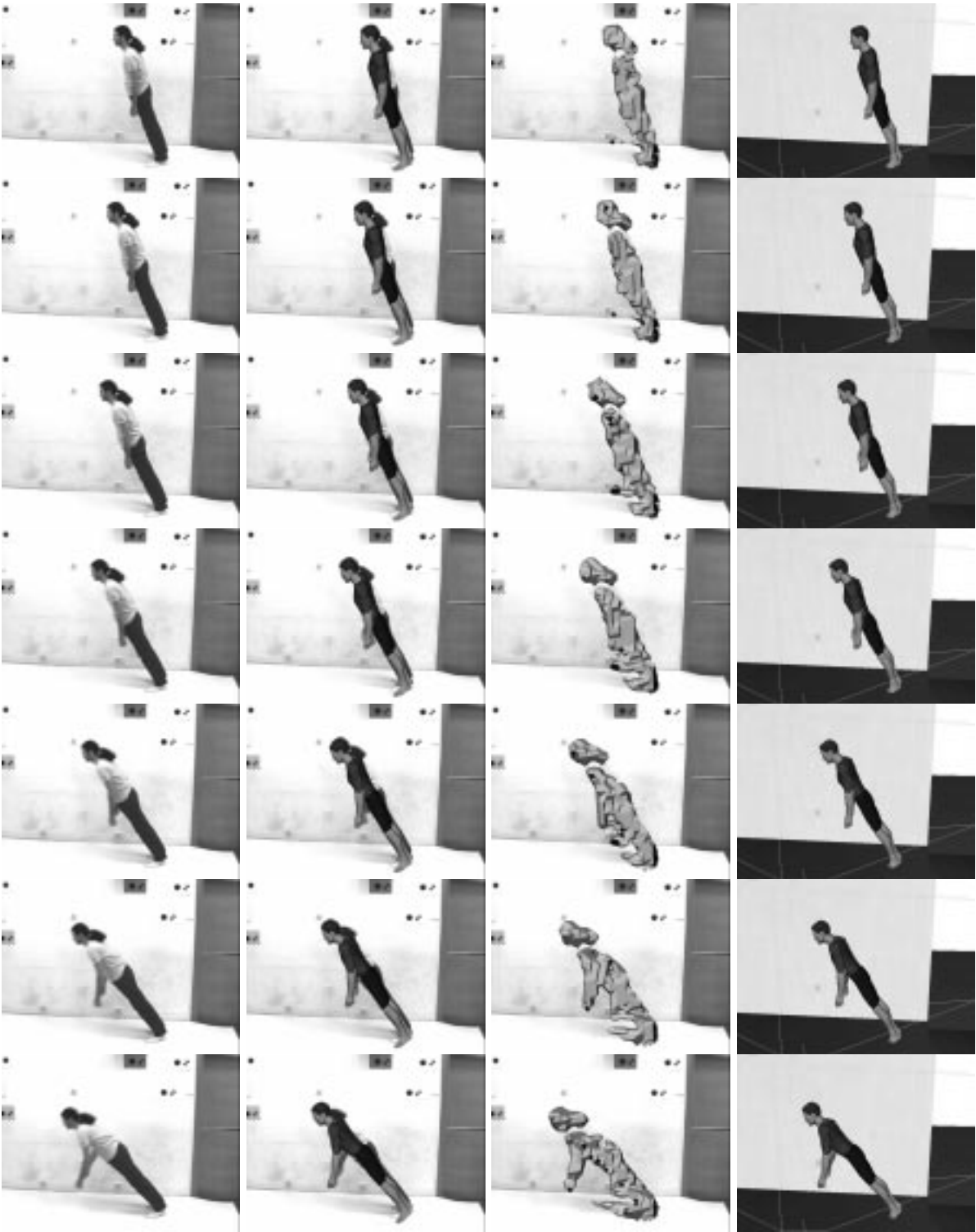
Position 3

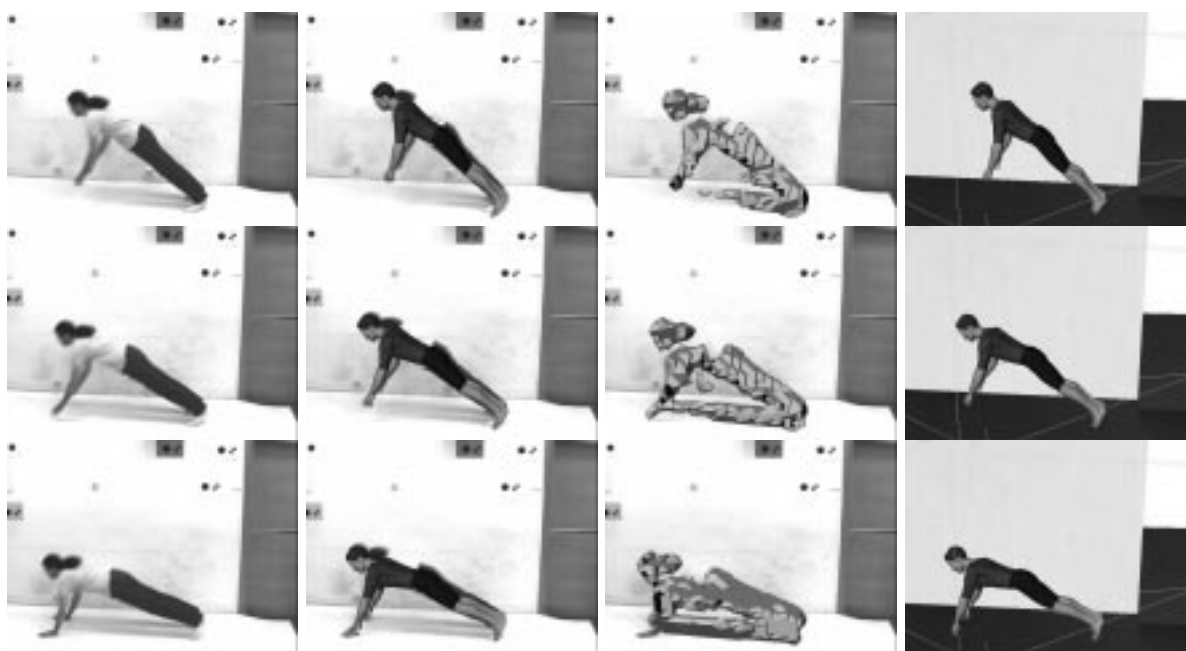


Position 4



Motion reconstruction (falling down) from frames 100 to 118, and the result of the application of the occupied volume compute algorithm.





Marching cubes reconstruction using colour images. Shadow problems are solved.





## 6.- Conclusions and future work

The system presented can analysed and generate a virtual human in H-Anim format with the same measurements than the subject, for later integrate it in the virtual world. For carry out this task, we capture the subject doing a motion and in a semiautomatic way the matching is carried out between the human and the humanoid. In this matching process the captured images is used as background image and humanoid is overlapped for verify the humanoid correct posture.

Once we have reconstructed the human motion can be erased the captured image and insert the humanoid in the virtual world doing the same motion. Now we are working with H-Anim specification humanoids and we are using also another specifications for the matching process.

In near future we plan to develop a high graph model including main terminal nodes (head, hands and feets), and the arcs will be the connections with a weight assigned value related with a fuzzy probability. Later by inverse kinematics and additional conditions try to search the others degree of freedom in the biomechanical tree chain.

In the other way the automatic process do the segmentation in real time with a low computing cost and can reconstruct the occupied volume. As application of this method we are developing a study for different kind of sports movements.

## 7.- References

- [1] J.M. Buades, A. Igelmo, F.J. Perales. "Modelos antropométricos a partir de secuencias de imágenes". CEIG 2000 X Congreso Español de Informática Gráfica. pp. 395-396, 2000.
- [2] J.M. Buades, R. Mas, F.J. Perales. "Matching a Human Walking Sequence with a VRML Syntehtic Model". AMDO 2000, First International Workshop. Palma de Mallorca, September 2000. pp 145-158.
- [3] C. Babski, D. Thalmann, "A Seamless Shape for H-ANIM compliant Bodies, Proc. VRML 99, ACM Press, pp 21-28, 1999.
- [4] F.J. Perales, J. Torres. "A system for human motion matching between synthetic and real images based on a biomechanical graphical model", IEEE Computer Society. Workshop on Motion of Non-Rigid and Articulated Objects, November 11-12, 1994, Austin Texas.
- [5] C. Yañiz, J. Rocha, F. Perales. "3D Part Recognition Method for Human Motion Analysis". CAPTECH'98 Modelling and Motion Capture Techniques for Virtual Environments. pp 41-55.
- [6] C. Wren, A. Azarbayejani, T. Darrell, A. Pentland. "Pfinder: Real-Time Tracking of the Human Body". IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 19, no 7, pp 780-785
- [7] C. Wren, A. Pentland. "Understanding Purposeful Human Motion", Submitted to ICCV 1999
- [8] D.M. Gravila, L.S. Davis. "3-D model-based tracking of humans in action: a multi-view approach", Computer Vision Laboratory, Proc. CVPR, pag 73-80, IEEE, 1996.
- [9] F. Perez, C. Koch. "Toward Color Image Segmentation in Analog VLSI: Algorithm and Hardware", International Journal of Computer Vision, 12:1, pp 17-42, 1994