

# Voxel Matching Reconstruction in Real Image Sequences of Human Avatars

## Submission Format

Jose Maria Buades Rubio  
Computer Graphics and Vision Group  
Department of Computer Science  
Universitat de les Illes Balears  
josemaria.buades@uib.es

Francisco Jose Perales Lopez  
Computer Graphics and Vision Group  
Department of Computer Science  
Universitat de les Illes Balears  
paco.perales@uib.e

---

### Abstract

*This article presents a part of a system that is used to analyse and synthesize human movement by means of a color segmentation and matching process to track and reconstruct the main aspects using a biomechanical model of a person. In particular, we explain the application of the Marching Cubes algorithm [Lorensen87] obtained from a set of voxel information. A color segmentation criterion is proposed. The segmented parts help us reduce the space search in the matching process. The main purpose of the system is to carry out a correspondence between this graphic model (primitives) and the person in movement in several real images (color and grey level).*

*The original voxel approximation is used to roughly fit the volume occupied by the person in the scene. The process is intended to be non-invasive and automatic, although it is currently used with the minimum manual intervention of the user. The system works in a controlled environment. 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 integration of people in virtual worlds, and others related to computer vision techniques, in order to recognise or track people in scenes. 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.*

### Keywords

*Human Motion Analysis, Background Substraction, Voxel Representation, Marching Cubes.*

---

## 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 and graphic 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, color segmentation, matching of entities, camera calibration, 3D reconstruction etc.), capturing the individual from an arbitrary number of color and grey level video cameras which enable us to obtain as much information as possible.

## 2. CAPTURING PROCESS

In the capturing process we have two possibilities; to use two synchronized color cameras, or four synchronized black and white ones. In this case we use only the two

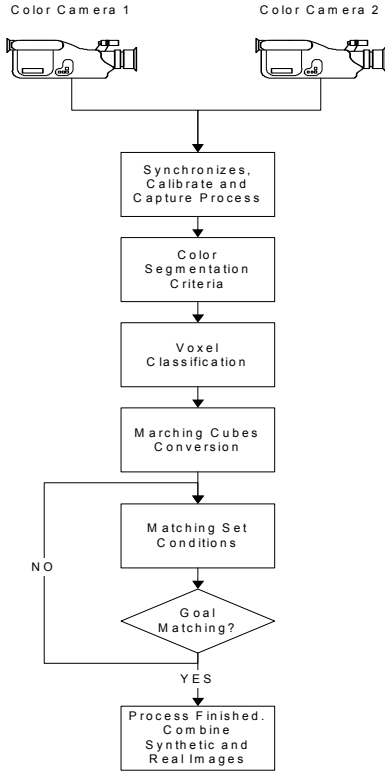
interlaced color cameras, because the voxel algorithm is based on color information.

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. Either way, the optics used do not introduce any appreciable distortion.

## 3. MATCHING CRITERIA

The matching process associates each articulation with a 2D point in the image and consequently a 3D projected primitive with a 2D region of pixels. 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. For the measurement, the person takes up four classic ergonomic postures which enable us to obtain the maximum information.



**Fig 1. Process Diagram, from capture process to final results**

On analysing the sequence, we are able to apply physical and temporal restrictions which help us carry out the matching whilst reducing the errors and the search space [Perales94]. This adjustment process is conditioned by a set of conditions which are optimised in each case and type of movement. The restrictions are:

**Inclusion Condition:** For all the segments of the 3D model projected onto the 2D sequences in all views, it is necessary that all segments or primitives are within the human body area obtained in the segmentation process. If the number of matched joints exceeds a predefined threshold, the condition is satisfactory.

**Distance Condition:** The segments of the body part should be centred in the boundary of segmented body part. This condition is used to adjust the position of the segment. Some errors are possible depending on the graphical primitive used (superellipsoids, generalised cylinders, etc.).

**Position Condition:** All the joints and segments should satisfy the physical conditions imposed by the biomechanical model. The algorithms proposed only search for the range within the physical angle constrains.

**Temporal Condition:** The position of the same joint or segment matched should be within a small distance from the position in the preceding frame. The distance for this condition is specified for a particular kind of movement.

**Collision condition:** There should be a one-to-one correspondence between nodes and segments and their positions in the 3D-space. If more than one primitive has the same 3D position, a collision is detected in the matching process. A match should not have collision.

**Variation Object-Model Condition:** We compute the variations between consecutive images and models to detect the joint variations. The idea is to reduce the number of segments updated in every iteration in the matching process.

This set of restrictions defines a level of matching. This matching process currently works in an automatic way for simple kinds of movements without multiple occlusions and no changes of topology.

This process is complex and time consuming so, we need to reduce the space search, using a voxel approximation where we suspect that the person is moving. This means that before using the set of conditions proposed, we need a segmentation process to detect occupied voxels and assign them to a part of the body.

For the time being, the automatic process detects the occupied volume, computed from the  $n$  different cameras, this is done in real time and at rate of 30 fps. To compute it we carry out the following steps:

The calibrated volume  $A \subset \mathbb{R}^3$  is partitioned in a subset of volumes  $X = \{A_1, A_2, A_3, \dots, A_N\} / \forall i A_i$  is a cube of size  $l$ ,  $\cup A_i \supseteq A$  and  $A_i \cap A_j = \emptyset$  if  $i \neq j$   
 $\forall i$  compute occupied( $A_i$ ).

### Algorithm 1. Compute first frame, all volume

This process is carried out for each captured frame, and for the total volume, but normally the volume the subject occupies is less 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 to be occupied, the algorithm computes the bounding box that contains the occupied volume, and only computes the bounding box for this subset volume. We use the following algorithm:

```

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

### Algorithm 2. Compute Next Frame

This allows us to restrict the studied volume to the zone of interest and to modify and move the bounding box according to the movements of the subject in relation with condition variation.

So far we have explained how to compute the volume in general, but we have not explained how to determine whether 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 a reference, with this reference image and the captured frame we carry out a subtraction and apply an adaptive threshold. We then have to discern two cases: color cameras and black and white cameras. In the b/w cameras the color space is one dimensional - the light intensity captured from the camera - therefore we have little useful information. In this case we only can apply one difference.

$$abs(frame(x,y) - background(x,y)) \geq threshold \quad (1)$$

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 color cameras, the color spaces are 3D, thus we have more information that we can use. After carrying out some tests with different color spaces we finally chose the HSI color space, rather than using YUV, RGB or nRGB. Now we have a typical threshold for each component H, S and I. This color space allows us to eliminate shadows caused by the subject and retrieve information about zones, which was not possible in b/w.

To select a voxel as changed from a camera we use an 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 component H.

$$\begin{aligned} abs(frame\_I(x,y) - background\_I(x,y)) &\geq threshold\_I \\ abs(frame\_S(x,y) - background\_S(x,y)) &\geq threshold\_S \\ abs(frame\_H(x,y) - background\_H(x,y)) &\geq threshold\_H \\ abs(frame\_H(x,y) - background\_H(x,y)) &\geq 360 - threshold\_H \end{aligned} \quad (2)$$

To achieve the best results, we previously smooth the captured and background images, which gives as a result the stabilization of the H-component and therefore divides the threshold of such component by two. For the I-component, we can allocate a high threshold and thus eliminate shadows without erasing parts of interest of the subject. The task of matching now focuses on determining out of the segmented images and the generated volume which parts are of interest.

Our proposed system is related to the work presented in [Arita00] but using more views, color and specific conditions. We also reduce the shape parameters and do not consider deformation. We are working to combine the process presented in the first part of this section with the last segmentation and voxel occupancy criteria. We know that recovering the structure of the body from the voxel representation is very difficult or may be impossible. We only use the voxel representation to fit the space and to help the matching process that uses simple predefined shapes. By selecting the end effectors such as

hands, feet and head and later applying inverse kinematics, we can achieve good results.

A few words are added here to explain briefly the 3D representation used. When we edit a humanoid, it is very helpful to have the person we are using as a 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. 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.

In the general automatic process, still in progress, 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.

#### 4. RESULTS

In this section, we present some examples of original sequence color images, segmented images, voxel overlapped approximation and marching cube cases. Results are illustrated in several figures. See figure 2.

#### 5. CONCLUSIONS AND FUTURE WORK

The system presented can analyse and match a segmented person with a biomechanical model in an automatic way. Finally, we can generate a virtual human in a compliant VRML format with the same measurements as the subject, for later integration in the virtual world. In order to carry out this task, we capture the subject doing a motion and the matching is carried out between the human and the humanoid. In this matching process the captured images are used as a background image and the humanoid is overlapped to verify the humanoid correct posture using the conditions considered.

Once we have reconstructed the human motion, the captured image can be erased and the humanoid inserted in the virtual world, going through the same motion.

As yet, the matching process proposed does not consider the facial features or the fingers of the person recorded. In the near future we plan to develop a high level graph model including main terminal nodes (head, hands and feet). Some important restrictions will be included when many changes of topology are presented. Also, to track hands and face we are testing an algorithm via particle filter [Isard98] which runs well with rigid regions and recovers partial occlusions.

On the other hand, the automatic process carries out the segmentation in real time with a low computing cost and can reconstruct the occupied volume. In the area of computer vision, we are trying to apply this system to recognise and track persons in a controlled environment.

#### 6. ACKNOWLEDGEMENTS

This project is subsidized by CICYT TIC2001-0931 and HUMODAN UE project IST Program

## 7. REFERENCES

- [Arita00] T. Nunomaki, S. Yonemoto, D. Arita, R. Taniguchi, "Multipart Non-Rigid Object Tracking Based on Time Model-Space Gradients", AMDO 2000. Palma de Mallorca, 2000. pp 72-82.
- [Gravila96] D.M. Gravila, L.S. Davis. "3-D model-based tracking of humans in action: a multi-view approach", CVPR 1996, pp 73-80.
- [HAnim] H-Anim 1.1 compliant VRML. <http://www.hanim.org>
- [Isard98] M. Isard, A. Blake "CONDENSATION-Conditional Density Propagation for Visual Tracking" International Journal of Computer Vision 29(1),5-28 (1998)
- [Lorensen87] William E. Lorensen and Harvey E. Cline, "Marching Cubes: A High Resolution 3D Surface Construction Algorithm", Computer Graphics. Proceedings of SIGGRAPH'87, Vol. 21, No 4, pp 163-169
- [Nielson91] Gregory M. Nielson, Bernd Hamann, "The Asymptotic Decider: Resolving the Ambiguity in Marching Cubes" Proceedings of Visualization'91 IEEE Computer Society Press pp 83-90
- [Perales94] F.J. Perales, J. Torres. "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, 1994, Texas.
- [Wren96] 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

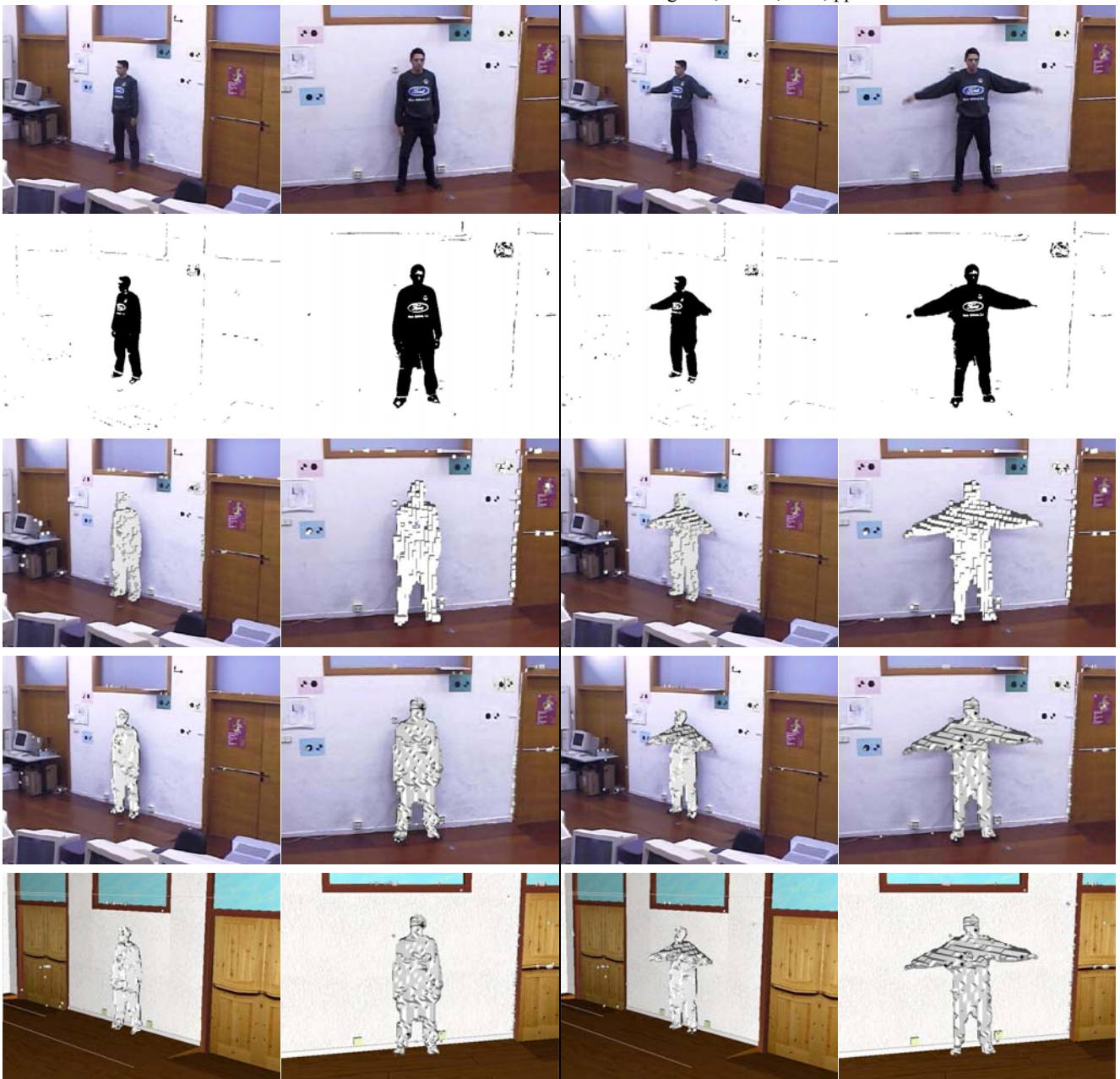


Fig 2. Two frames, left frame number 190, right frame 250 from two color cameras. Top to bottom: real image, segmentation, voxel representation, marching cubes rendered image, and virtual world mixed marching cubes