Human model and pose Reconstruction from Multi-views

Michoud Brice
LIRIS-CNRS UMR 5205
Université de Lyon 1
Email: b-mich02@liris.cnrs.fr

Guillou Erwan
LIRIS-CNRS UMR 5205
Université de Lyon 1
Email: eguillou@liris.cnrs.fr

Bouakaz Saïda
LIRIS-CNRS UMR 5205
Université de Lyon 1
Email: bouakaz@liris.cnrs.fr

Abstract—We present an approach for automatic acquisition of human body model and pose estimation from multiple synchronized video streams. To represent the human shape we adopt an animated articulated skeleton-based model. A skeleton is a set of segments joined by articulations, which are generally rotations, to which are added (anatomical) constraints of variations of angles. First, we assume the foreground silhouette is extracted from each video frame. These silhouettes are then used as the input for the extraction of the body part, localization of skeleton joints and shape model representation. The originality of our approach is the extraction of a real skeleton of animation using silhouettes. The addition of a geometrical volume around the skeleton through geometrical primitives allows to obtain 3D geometrical representation of the shape of the avatar. The reprojection of the texture (color, texture) extracted from every view of human body, allows to give a realistic appearance to the reconstructed avatar. The approach which we propose is simple. In opposite of most of the existing methods, it imposes no constraint except the fact that the filmed person has to be in stable pose. The results obtained in the real cases allow to validate the approach.

I. INTRODUCTION

Due to the large potential of applications in the field of video conference, augmented reality, telepresence, human machine interaction, etc, recovering the human body model and pose estimation received significant interests. Given the dimension of the work, it is difficult to address a complete state of the art, so we are focusing on particular applications with related interests to our work.

These citations are by no mean complete and admit further classification, descriptive details, and enlargement of scope of application. Several methods had been proposed for reconstructing three dimensional models of human shape and motion tracking from one [1] or multiple views [2], [3], [4]. Many motion capture systems are developed, several surveys and recent works interested in this task are addressed [5], [6], [7]. Broadly speaking, the proposed methods can be divided into two categories.

The first category is interested in detail model reconstruction [8], [9], [10] and includes both techniques developed in computer graphics and image synthesis domain [11]. These techniques care, in general about giving high degree of realism in appearance and rendering or/and require expensive process. Typically, model estimation and localization is turned as multi dimensional expensive search [12], [13], [14]. Indeed, correspondences between the extracted part of real images and model, is obtained either by some stereo matching techniques [15], [5], [16] or by minimizing some cost function which is defined in terms of correlation between model prediction and real data [1], [17]. In [18] a force based tracking allows to align the out line of the model to the contour extracted from the frame. Although the acquisition frame and some pre-processing steps are done in real time, the model fitting and pose estimation or analysis step are, in most cases, an expensive process and they are done off-line.

The second category involves real time, however it is intended to applications which require few details. One popular approach is voxel based reconstruction [19], [20]. The first attempt in using voxel representation to estimate body pose has been proposed by Cheung et al. [21]. It involves several steps, including silhouette extraction for images, visual hull estimation to represent the shape of the human body from the silhouettes, voxel reconstruction by projecting the silhouettes on an plane depth, and use of the marching cube algorithm to produce a surface from the voxel data. Whereas some papers argue the feasibility of the method in real-time, the realism of the visual rendering is still below the expected requirements. Many other frameworks are proposed in real time context. Almost all algorithms related to this category assume that a generic 3D body model is given. As it is pointed above, the estimating body posture and/or tracking step is done by adjusting each part of the generic model to real images/silhouettes [7]. Although these approaches work well, they assume that the body model of the tracked person is placed close to the true position in the beginning, and the size of different parts (e.g. anthropometric measures) of body are given or hand adjusted. These algorithms then estimate the model parameters during the time in order to express the motion of the person.

We argue that for full automated body model reconstruction in addition of posture tracking, challenge is to solve model acquisition problem and the initial estimation of the initial pose.

By model acquisition we mean to estimate the specific anthro-
pometric measures (e.g. shape and size of the person). Solving the initial posture consists in the placement and configuration of different parts of the model, according the posture of the body in the beginning of the motion capture process.

In this paper, we present an approach for an automated 3D model body reconstruction that includes both the model acquisition and posture estimation.

The paper proceeds with a brief overview of our approach in section II followed by preliminaries about skeleton representation and joint extraction. Section III and section IV detail descriptions of each step of extracting joints from the silhouettes and reconstructing the body animation skeleton; details are given on how to recognize right limbs and left ones. The section V presents a body hull reconstruction. This body hull is represented by a set of geometrical primitives. Finally, experimental results, conclusion and future work are given in section VI and section VII.

II. OVERVIEW AND PRELIMINARIES

Generally, in computer graphics domain, human body is represented by an animation skeleton. Geometrical primitives are attached to the skeleton and give the appearance of human body 3D shape. Hence estimating the pose of a person can be seen as the process that computes the positions of the joints (as described in fig. 3(a)) in the human skeleton in respect to the binary mask of silhouettes acquired from each camera. For recovering body posture and 3D model reconstruction, first we introduce and extract a Condensed body Representation (CR). The CR representation is a set of feature (representative) points that performed using the principle of cross section that given by intersection of perspective projection silhouettes on selected plane. Analysing CR representation allows estimating several joints of skeleton. Therefore, we process this skeleton by adding geometric constraints that characterize the human body such as the standard length ratio of the limbs. Given the skeleton and perspective projection silhouettes of human body, geometrical primitives are fitted to the data for subsequent shape representation.

One difficult problem is to generate body silhouette by background subtraction and to remove the shadows. In our case we use the method described in [21]. Our method works on the bounding box $B$ of the 3D space occupied by this person. This box can be estimated from the frustum of the cameras.

A. Slices

Let $\pi$ be a plane defined by a point $p$ and a normal vector $\vec{n}$. We will call slice, the 2D image describing the intersection between $\pi$ and the 3D shape of the person being captured by cameras.

As we do not know the shape of the person, slices will be computed using projective texture mapping[22] method with the binary masks as projected textures. The projection is done on a quad defined over the $\pi$ plane and all projected textures are combined using a logical AND operator (see fig. 1).

Notice that the set of computed slices can be seen as an estimation of the voxel reconstruction ($V_{\alpha\xi}$) of the person.

Let $S_j$ the computed slice over plane $\pi_i(p_i, \vec{n}_i)$, $S_{j+1}$ will denote the following slice of $\pi_i$ in the direction described by the normal vector $\vec{n}_i$. The skeleton joints will be extracted from representative points selected in the set of slices computed over the bounding box $B$ (see fig. 2(a)). Those representative points are linked together when extracted from two successive slices.

B. Condensed representation

Let $\Psi$ be the set of $J$ connex components found in slice $S$ and $\zeta_j \in \Psi$, one of those connex components. We will use $R_{p_j} = \{G_j, A_j, F_j, E_j\}$ as the representative point of $\zeta_j$, where:

- $G_j$ describe the 3D position of the center of gravity computed over $\zeta_j$.
- $A_j$ is the area of $\zeta_j$.
- $F_j$ is the set of rim points of $\zeta_j$.
- $E_j$ defines the ellipse of inertia computed over $\zeta_j$.

We introduce the concept of condensed representation (CR) of a set of slices $S$, the set of all representative points $R_{p_j}$ computed over all slices of $S$. Over this set of representative points, we define an oriented (in the direction of the normal vector $\vec{n}_i$ of the slices) linkage (see fig. 2(b)) described by:

two representative points $R_{p_i}$ and $R_{p_j}$ are linked if:

1) they belong to two successive slices;

2) they are connected by an edge $\epsilon$.

3) $\epsilon$ belongs to a slice $S_k$.

4) $R_{p_i}$ and $R_{p_j}$ belong to $S_k$.

5) $S_k$ is the slice that minimizes the distance between $R_{p_i}$ and $R_{p_j}$.

Figure 2: Slices and condensed representation; (a) bounding box and successive slices; (b) condensed representation; (c) $\alpha$-traversal of the CR described in (b) ($\alpha=0.3$).
2) the connex components associated to Rp_i and Rp_j are connex in the voxel representation Voxel.

Notice that a representative point can be origin (or extremity) of many individual links. In order to reduce the set of the representative points traversed at the time of joint extraction, we should treat only the significant connections of the global linkage of the CR. This is done using the concept of α-representativeness : a representative point Rp_i is α-representative if the area A_i of the associated connex component verify equation 1.

\[
\sum_{j=1}^{\infty} A_j(\zeta_i, \zeta_j) \in \Phi^2
\]

By extension of this concept, we define the α-traversal as the oriented traversal of all α-representative points in the CR.

C. Joint extraction

According to the assumption previously defined (the acquired person is upright in steady balance), the most constrained (in terms of DOF\(^1\)) parts of its body are the legs. Hence, the first step of the method consists in extracting the lower joints (waist, hips, knees, ankles and feet). Knowing the 3D position of these joints, the second step extracts the upper joints (head, neck, shoulders, elbows, wrists and hands). This sequence of extraction is denoted in fig. 3(b) for each group of joint.

![Fig. 3. (a) labelled animation skeleton; (b) sequence of the joints extraction.](image)

III. EXTRACTION OF LOWER JOINTS

Our method is based on a prediction / correction strategy. It first extracts an estimate for the 3D position of waist and hips joints, uses these estimates to computes 3D position of the joints in the legs and then corrects waist and hips position using measurements in the legs.

A. Waist and hips estimates

The process that looks for the waist joint in the CR is equivalent to the one that seek the configuration described in fig. 4: the representative point where the legs local linkage connects to the bust local linkage. The connection point will be used as an estimate of the waist joint (noted \(R_{\text{waist}}\)). To find this point, we use an α-traversal on the global linkage of the CR from its lowest points (\(R_{\text{foot}}\)) in fig. 4. This traversal will stop with success on the first detected confluence. From the experiments, using the value \(\alpha = 0.3\) (see fig. 2(c)) gives good results avoiding slice computation artefacts.

The two points backward linked to the waist estimate are used as an estimate for the position of the hip joints (\(J_{\text{hip1}}\) and \(J_{\text{hip2}}\) in fig. 4). We will show later how alleviate ambiguity between the left-hand side and the right-hand side.

B. Generic estimates of limb’s joints

We propose a generic method that estimates the position of the joints describing a limb. Knowing two extreme points A and B on the limb, we construct a local CR \(L_{\text{limb}}\) over the bounding box of the limb.

This CR is oriented along the vector \(\overrightarrow{AB}\). A is the connection point between the limb and the bust, B is a point on the other side of the limb (on the foot or the hand, depending on the limb).

![Fig. 5. Steps of the generic estimation of the limb's joints.](image)

First, we choose an estimate for the extreme point \(R_{p3}\) of the limb (hand or foot) as the nearest representative point (in the local CR) from point B. Next, we find an initial estimate for \(R_{p1}\) (estimate for the knee or the elbow) as the point of \(L_{\text{limb}}\) equidistant from \(R_{p0}\) (nearest representative point from A) and \(R_{p3}\) (see fig. 5(a)), as defined in equation 2.

\[
R_{p0}R_{p1} = R_{p1}R_{p3}
\]
Then we estimate $\mathbf{R}_2$ (estimate for the ankle or the wrist) as the point of $L_{limb}$ the farthest (using an euclidian distance from point to line) from the line $(\mathbf{R}_1, \mathbf{R}_3)$ (see fig. 5(b)) using equation 3 where $\mathbf{R}_p$ belongs to the sub-linkage of $L_{limb}$ defined between points $\mathbf{R}_1$ and $\mathbf{R}_3$.

\[ d_c(\mathbf{R}_2, (\mathbf{R}_1, \mathbf{R}_3)) = \max d_c(\mathbf{R}_p, (\mathbf{R}_1, \mathbf{R}_3)) \quad (3) \]

Lastly, we refine the position of $\mathbf{R}_1$ as the point of $L_{limb}$ the farthest from the line $(\mathbf{R}_0, \mathbf{R}_2)$ (see fig. 5(c)), using equation 3 applied to the sub-linkage of $L_{limb}$ between points $\mathbf{R}_0$ and $\mathbf{R}_2$.

C. Leg’s joints estimation

Leg’s joints estimation is done using our generic method. Point $A$ (resp. $B$) is associated with $\mathbf{J}_{hip}$ (resp. $\mathbf{R}_foot$) the hip joint previously estimated (resp. foot joint estimated as the first representative point of the local linkage connected to $\mathbf{J}_{hip}$). From this method, we obtain $\mathbf{J}_{ankle}$ from $\mathbf{R}_2$, $\mathbf{J}_{knee}$ from $\mathbf{R}_1$ and $\mathbf{R}_foot$ from $\mathbf{R}_0$. Notice that the point $\mathbf{R}_0$ replace the initial estimation for $\mathbf{J}_{hip}$.

Knowing the position of the joints in a limb, we can correct less accurate estimations using statistical measurements made a set of representative persons. For example, the length of a leg is approximately equal to the half-height of the person [23].

Leg’s joints estimation are more accurate when the leg is bent. In order to have the same precision on the positions of the joints of the two legs, we use length of the rigid parts of the most bent leg to translate position of the knee joint on the other leg.

D. Waist and hips correction

The estimates for waist and hips joints are localized on the visible separation between the two legs. Although these estimations are sensitive to the clothes worn by the acquired person, the orientation of the thighs given by $\mathbf{J}_{knee}$, $\mathbf{J}_{hip}$ are useful informations. Using this orientation, $\ell_c$ the length of the calf and $\ell_f$ the length of the foot, we can correct the position of $\mathbf{J}_{hip}$ using equation 4.

\[ \mathbf{J}_{hip} = \mathbf{J}_{knee} + \frac{\mathbf{J}_{knee} - \mathbf{J}_{hip}}{\mathbf{J}_{knee} - \mathbf{J}_{hip}} (\ell_c + \ell_f) \quad (4) \]

Then $\mathbf{J}_{waist}$ is placed as the center of the segment $[\mathbf{J}_{hip1}, \mathbf{J}_{hip2}]$.

E. lefthand/righthand ambiguity

Now that an estimate for the positions of lower joints is known, we can alleviate the ambiguity between lefthand side and righthand side of the acquired person. Let $T$ be the trihedron having joint $\mathbf{J}_{ankle}$ as top point and a basis defined by joints $\mathbf{J}_{hip1}$, $\mathbf{J}_{foot1}$ and $\mathbf{J}_{ankle2}$ (see fig. 6). If $T$ is direct, then joints with an index value of 1 (resp. 2) define the joints on the right (resp. left) side. Notice that this method allows us to compute $\mathbf{V}_{front}$, the vector describing the front side of the acquired person.

IV. EXTRACTION OF UPPER JOINTS

Now that we have shown how to extract an estimate for the position of each of the lower joints, we can deal with the upper part of the human body: from waist to head, including arms. In the first part of this section we deal with the bust and the head. Hence, representative points in the CR corresponding to arms and legs are not useful (see fig. 7(a)). It is thus necessary to filter the points not having important significance. Accordingly, we define $L_{bust}$ as the sub-linkage from $\mathbf{J}_{waist}$ containing the representative points having a maximum area on each slice (see fig. 7(b)). Notice that $L_{bust}$ links only one representative point per slice.

The first step in the upper joints estimation process consists in finding the position of the head. Knowing this information, it proceeds with shoulders estimation then looks for joints in the arms.

A. Neck and head estimation

Let $\mathbf{J}_{head}$ be the upper representative point in $L_{bust}$, it corresponds to the highest observable point on the head of the acquired person. The method seeks the representative point $\mathbf{J}_{neck}$, intersection point between the head and the bust (i.e a point on the neck, see fig. 7(b)). The set $\mathbb{I}_{neck}$ of applicant representative points for $\mathbf{J}_{neck}$ is defined using equation 5 where $\mathbf{R}_p$ is a representative point in $L_{bust}$ and $\ell_h$ is the
euclidean distance from $J_{\text{waist}}$ to $J_{\text{head}}$ (see fig. 8(a)).

$$I_{\text{neck}} = \{ R_p, \frac{2\ell_h}{3} < J_{\text{waist}} R_p < \ell_h \}$$  \hspace{1cm} (5)

Let $P_{\text{pyr}}$ be the regular pyramid oriented along $J_{\text{head}} J_{\text{waist}}$, having $J_{\text{head}}$ as top point and a square base (centered on $J_{\text{waist}}$, side length $\ell_h$). Neck joint $J_{\text{neck}}$ will be chosen as the representative point $R_p_{\text{neck}}$ verifying equation 6, where $R_p_i$ is a representative point in $I_{\text{neck}}$ and $\pi_i$ is the plane over which slice containing $R_p_i$ has been computed (see fig. 8(b)).

$$\frac{A_{\text{neck}}}{\text{Area}(\pi_{\text{neck}} \cap P_{\text{pyr}})} = \min \frac{A_i}{\text{Area}(\pi_i \cap P_{\text{pyr}})}$$  \hspace{1cm} (6)

![Diagram](image)

Fig. 8. (a) 2D position of $J_{\text{neck}}$ and $J_{\text{head}}$; (b) curves describing the search for $J_{\text{neck}}$.

B. Shoulders estimation

Using the same statistical measurements as for limb correction (see section III-C) or the perfect human ratios given by Da Vinci, we have some hypothesis on the shoulders position: (1) the shoulders are located at two thirds of the trunk; (2) the width of the shoulders is equal to the quarter of the size of the acquired person.

Hence, we choose the main\(^2\) connex component $\zeta$ on a slice computed over the plane parallel to the reference plane and containing the point $E$ defined by equation 7.

$$E \in [J_{\text{waist}} J_{\text{head}}] \text{ and } J_{\text{waist}} E = \frac{2}{3} J_{\text{waist}} J_{\text{head}}$$  \hspace{1cm} (7)

The representative point $R_p$ associated with $\zeta$ is used as an estimate for joint $J_s$, connection between bust and shoulders. Shoulders’ joints $J_{\text{shoulder}}_l (J_{s_l})$ and $J_{\text{shoulder}}_r (J_{s_r})$ are then estimated as points along the major axis of inertia of $\zeta$, using equation 8 where $\ell_H$ is the estimated size of the acquired person.

$$J_{s_l} J_{s_r} = \frac{\ell_H}{5} \text{ and } J_s = \frac{J_{s_l} + J_{s_r}}{2}$$  \hspace{1cm} (8)

\(^2\)the one with maximum area

C. Arm’s joints estimation

Arm’s joints estimation is done using our generic limb estimation method (see sec. III-B). Point $A$ is associated with $J_{\text{shoulder}}$, we have to find $P_{\text{arm}}$ a point at the extremity of the arm.

The first step consists in finding a set of applicant points. Let $\pi$ be the plane defined by $J_{\text{waist}}$ and $V_{\text{front}} \wedge J_{\text{waist}} J_{\text{head}}$. From global CR deprived of legs and head representative points ($\overline{CR}$), we build two sets $S_l$ and $S_r$ of applicant points. Let $P$ be a rim point in $\overline{CR}$. $P$ belongs to $S_r$ if it is above $\pi$ and to $S_l$ otherwise (see fig. 9(a)).

![Diagram](image)

Fig. 9. (a) sets of applicant points for arms; (b) search spaces for $P_{\text{a}}$.

From now, we’ll explain the process of finding the joints in one arm, hence $l$ and $r$ indices won’t be used. Let $P_{\text{arm}}$ be a point in $S$ defined using equation 9.

$$d_e (P_{\text{arm}}, (J_{\text{neck}}, J_{\text{waist}})) = \max_{P \in S} d_e (P, (J_{\text{neck}}, J_{\text{waist}}))$$  \hspace{1cm} (9)

Let $I([ab])$ be the ratio of segment $[ab]$ that felt inside the 3D shape of observed human. This ratio is computed using the representative points in the CR.

- $P_{\text{arm}}$ is an estimate for $J_{\text{hand}}$ if it verifies equation 10 or 11.

$$P_{\text{arm}} J_{\text{shoulder}} \simeq J_{\text{hip}} J_{\text{knee}} + J_{\text{knee}} J_{\text{foot}}$$

$$I([J_{\text{shoulder}} P_{\text{arm}}]) < 0.9$$

- Otherwise, $P_{\text{arm}}$ is an estimate for the elbow. The joint $J_{\text{hand}}$ is searched in two search spaces (see fig. 9(b)) depending on how the forearm is bent in respect to the brachium. This is done by finding a point $P_a$ located in the forearm.

First, we look in search space $\text{search}_1$ (as defined in fig. 10(a)). Let $P_c$ be a point on line $(P_{\text{arm}} J_{\text{shoulder}})$ defined by equation 12 where $\epsilon$ is the minimal distance between two points in the CR.

$$P_c = J_S + (1 + \epsilon) J_S P_{\text{arm}}$$

Let $S_{\text{arm}}$ be the slice computed over the plane defined by the point $P_c$ and its normal vector $P_{\text{arm}} J_{\text{shoulder}}$. Two cases can then occur:

1) either there is a connex component $\zeta_i$ on $S_{\text{arm}}$ whose center of gravity $G_i$ verifies equation 13, then...
the representative point associated with $\zeta_i$ is used as an estimated for $P_a$.

$$I([G_i P_{arm}]) > 0.9 \text{ and } I([G_i J_{shoulder}]) < 0.7$$

(13)

2) otherwise, we have to find $P_a$ in search space search2 (as defined in fig. 10(b)). Computing successive slices $S_i$ from $P_{arm}$ in the direction described by vector $P_{arm} J_{shoulder}$. We search for the first slice $S_j$ having a second connex component $\zeta_{j,1}$ not aligned with the brachium (the first connex component $\zeta_{j,0}$ represents the brachium itself) whose center of gravity $G_{j,1}$ verifies equation 14. Then $G_{j,1}$ is then used as an estimate for $P_a$.

$$I([G_{j,1} P_{arm}]) > 0.9 \text{ and } I([G_{j,1} J_{shoulder}]) < 0.7$$

(14)

At this step of the method, we have two special points $P_{arm}$ and $P_a$ defining the orientation of the forearm. A point $P_{hand}$ on the hand is then estimated using the fact that brachium and forearm have equivalent length as defined in equation 15.

$$P_{hand} = P_{arm} + \frac{P_{arm} J_{shoulder}}{P_{arm} P_a} P_{arm} P_a$$

(15)

Inside joints of the arm ($J_{wrist}$ and $J_{elbow}$) are estimated using the generic limb method (see sec. III-B). Point A (resp. B) is associated with $J_{shoulder}$ (resp. $P_{hand}$). To sum up, we obtain $J_{elbow}$ from $R_{p1}$, $J_{wrist}$ from $R_{p2}$ and $J_{hand}$ from $R_{p3}$.

V. BODY HULL ESTIMATION

We have described the first step of our method which allows us to estimate the position of the human body joints. We’ll now introduce the second step which deals with the body hull reconstruction. This body hull is described by a set of 3D primitives which will be used when rendering the person in a virtual environment.

A. Body hull representation

In commonly used methods for reconstruction (voxels, visual hulls etc.), the obtained model do not have any semantic information: it’s very hard to know what part of the body is represented by any polygon (or voxel). Moreover, the set of polygons will change when performing motion analysis over a video sequence. Hence, it’s impossible to ensure that a known polygon from the initial reconstruction will always exists, and will be associated to the same part of the human body, over the sequence.

![Fig. 10. (a) first step of search for $P_a$; (b) second step of search for $P_a$](image)

![Fig. 11. 2D representation of the human body (visualized over a generic skeleton).](image)

The skeleton estimated in section III and IV describes a hierarchy of joints. Those joints are linked and links represent rigid parts of the human body (foots, forearms, ...). Our method estimates the 3D envelope of the subject by representing volumes which describe these rigid parts; Volumes which will be defined by a set of generalized cones (as shown in fig. 11(a), figures 11(b) to 11(e) describing the main parts of the human body).

B. Cone bases estimation

The base step of the method is the estimation of the cone base. When representing a shape with a generalized cone, we compute a slice for each base of the cone. On these slices, we choose one connex component (the one for which the center of gravity is the nearest to the center of the image). The cone base will be defined as the ellipse of inertia of the connex component centred on the center of gravity (see fig. 12).

![Fig. 12. estimated ellipse on the image of a slice.](image)

C. Body hull reconstruction

Our body hull estimation method can be divided in the estimation of three main parts of the human body:

- **The bust** is described by 4 generalized cones (see fig. 11(b)) whose bases are defined over five slices. These slices are computed over five planes parallel to the reference plane previously defined. The planes are placed between $J_{waist}$ and $J_{neck}$ as described in fig. 13(a).
- **head** is reconstructed with five cones (see fig. 11(c)) whose bases are defined over six equally spaced between $J_{neck}$ and $J_{head}$. As for the bust, these planes are parallel to the reference plane (see fig. 13(b)).
• arms and legs are built in the same way, using a set of generalized cone for each rigid part between two joints (see fig. 11(d) and 11(e)). In the skeleton model, they are described by four joints $J_i$ and linked to the bust with a joint $J$ (see fig. 13(c)). Cone bases are defined by slices computed over planes whose normal vectors are defined by the following equations:

\[ J_0 J_1 \] at $J_0$

\[ (J_0 J_1 + J_1 J_2)/2 \] at $J_1$

\[ (J_1 J_2 + J_2 J_3)/2 \] at $J_2$

\[ J_2 J_3 \] at $J_3$

\[ J_{i-1} J_i \] at $b_i$

Points $b_i$ are the centers of the rigid part defined by $J_{i-1}$ and $J_i$.

VI. RESULTS

The proposed method has been tested on several datasets (as shown in fig 14). The computation has been made on a Pentium IV 2.8 Ghz, with 512 Mo of RAM and a NVIDIA GeForce FX 5900 graphic card.

A. Body posture estimation

Posture estimations are shown in fig 15 where the second column shows the projection of the estimated skeleton on the binary mask acquired from one camera. As we can see, posture estimation gives suitable results on the orientation of the arms even if the hands are in contact of the body, as shown for the third dataset.

<table>
<thead>
<tr>
<th>datasets</th>
<th>average variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>length</td>
<td>(a)</td>
</tr>
<tr>
<td>thigh</td>
<td>48.29 $\pm$ 0.01</td>
</tr>
<tr>
<td>calf</td>
<td>32.83 $\pm$ 0.01</td>
</tr>
<tr>
<td>foot</td>
<td>20.68 $\pm$ 0.71</td>
</tr>
<tr>
<td>leg</td>
<td>101.80 $\pm$ 0.71</td>
</tr>
<tr>
<td>brachium</td>
<td>38.35 $\pm$ 0.65</td>
</tr>
<tr>
<td>forearm</td>
<td>26.12 $\pm$ 1.47</td>
</tr>
<tr>
<td>hand</td>
<td>13.52 $\pm$ 1.01</td>
</tr>
<tr>
<td>arm</td>
<td>78.00 $\pm$ 1.12</td>
</tr>
</tbody>
</table>

All those datasets are acquired from the images of a same person. To study the correctness of the estimation, we compare results (in terms of body lengths) between the different datasets. Those comparison are shown in tab I.

As we can see, our method gives good results on the body length with an average variation below 2%. The measures in the legs are equally estimated, differences between estimation depending on the position of the knee joint inside the human body; better results are given when the estimated position of this joint is centered inside the 3D volume of the knee. The maximum variation is achieved with the estimation of the arms. This is due to the number of DOF$^3$ of the arms; as a matter of fact, there is no constraint on the position of the arms in respect to the other parts of the body.

---

$^3$degrees of freedom
Fig. 16. Body hull reconstruction (from dataset (a) in fig 14): (a) using voxels (voxelspace resolution set to 256^3), computation time 580 ms; (b) using our method, computation time 44 ms.

(a) from fig 16) and the original binary masks (dataset (a) in fig 14). Those differences, which do not appear for the voxel reconstruction, are mostly due to the small number of generalized cone used.

One of the advantages of our method, in respect to voxel reconstruction, is shown in fig 17. Our method is based on well-known primitives hence artefacts shown by voxel reconstruction do not arise in our method. Moreover, from its nature, our method makes it possible to associate semantic informations to the estimated geometry and thus is able to tell which cone corresponds to any parts of the body.

Fig. 17. Virtual body rendering from novel viewpoint.

VII. CONCLUSION ET PERSPECTIVES

In this paper, we have proposed a method to recover a human body posture in model context: we can recover information about body skeleton as joint locations limb size. We have outlined techniques for extracting a complete animation skeleton model. We have presented a process based on simple geometrical primitives for reconstruction the 3D model of a person. The recovered shape and skeleton parameters can be used to create a realistic animation and/or to track human motion. Our model and pose acquisition approach is automatic, self-adaptive, and does not require any special movements by the filmed person.

Our representation, while having a fairly low dimension, nevertheless has a rich enough components structure to yield good performance at a low computational cost. At present, the limitation comes from the small size of the database we use. It should be completed; this should allow us to precisely undertake wider range of styles, and extend our work to more complicate cases. In particular, until now, the cases that we deal contain only one person. One of the extensions would be to acquire sequences containing several persons in the same time and the same place.

ACKNOWLEDGMENT

The work reported here was funded in part by the French National Research Agency. The author would like to thank members of Cyber II Consortium for their help.

REFERENCES