November 1994

Active Motion-Based Segmentation of Human Body Outlines

Ioannis A. Kakadiaris
University of Pennsylvania

Dimitris Metaxas
University of Pennsylvania

Ruzena Bajcsy
University of Pennsylvania

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-94-58.

Active Motion-Based Segmentation of Human Body Outlines

Abstract

We present an integrated approach towards the segmentation and shape estimation of human body outlines. Initially, we assume that the human body consists of a single part, and we fit a deformable model to the given data using our physics-based shape and motion estimation framework. As an actor attains different postures, new protrusions emerge on the outline. We model these changes in the shape using a new representation scheme consisting of a parametric composition of deformable models. This representation allows us to identify the underlying human parts that gradually become visible, by monitoring the evolution of shape and motion parameters of the composed models. Based on these parameters, their joint locations are identified. Our algorithm is applied iteratively over subsequent frames until all moving parts are identified. We demonstrate our technique in a series of experiments with very encouraging results.

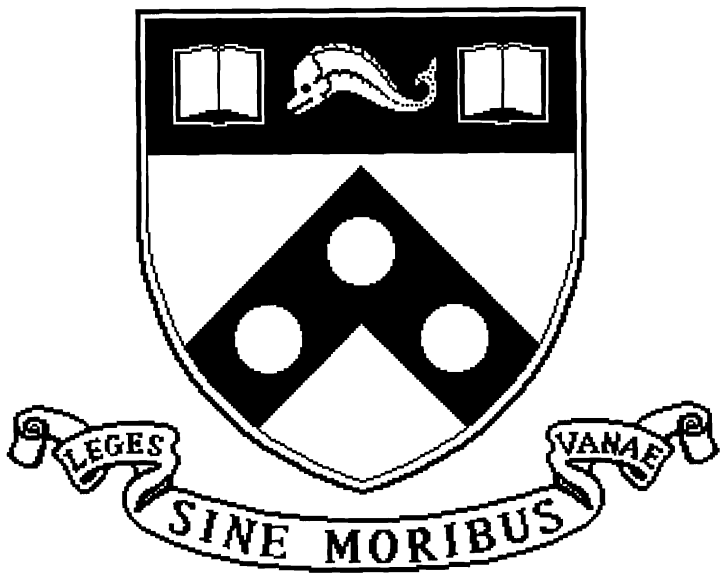
Comments

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-94-58.

Active Motion-Based Segmentation of Human Body Outlines

MS-CIS-94-58
GRASP LAB 382

Ioannis A. Kakadiaris
Dimitri Metaxas
Ruzena Bajcsy



University of Pennsylvania
School of Engineering and Applied Science
Computer and Information Science Department
Philadelphia, PA 19104-6389

November 1994

Active Motion-Based Segmentation of Human Body Outlines

Ioannis A. Kakadiaris, Dimitri Metaxas and Ruzena Bajcsy
GRASP Laboratory
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104

Abstract

We present an integrated approach towards the segmentation and shape estimation of human body outlines. Initially, we assume that the human body consists of a single part, and we fit a deformable model to the given data using our physics-based shape and motion estimation framework. As an actor attains different postures, new protrusions emerge on the outline. We model these changes in the shape using a new representation scheme consisting of a parametric composition of deformable models. This representation allows us to identify the underlying human parts that gradually become visible, by monitoring the evolution of shape and motion parameters of the composed models. Based on these parameters, their joint locations are identified. Our algorithm is applied iteratively over subsequent frames until all moving parts are identified. We demonstrate our technique in a series of experiments with very encouraging results.

1 Introduction

The task of human motion analysis can be decomposed into the following subtasks: (1) estimation of the shape and motion parameters of the parts of a human body, and (2) recognition of the motion performed by the actor. The estimation subtask is important in applications such as anthropometry, human factors design, ergonomics, performance measurement of both athletes and patients with psychomotor disabilities and virtual reality. On the other hand, the recognition subtask is important in applications such as human computer interaction.

The human body consists of non-rigid articulated parts whose motion depends on joints with various shapes and on intricate muscle actions. To recover the degrees of freedom associated with the shape and motion of a moving human body, many researchers either use a model-based approach [10, 4, 1, 7, 8, 15, 14] or employ certain assumptions [13, 18, 3, 17, 2, 19, 6]. Most of these techniques assume models that can only approximate the human body (e.g., generalized cylinders) and cannot adapt to different body sizes, since they are not deformable. To overcome this problem other researchers assume prior segmentation of the given data into parts [12] and then fit deformable models that can adapt to data from humans of different sizes [11, 9]. However, the process of segmentation and the process of shape and motion estimation are decoupled leading to possible lack of robustness and inaccuracies. Moreover, no techniques exist that acquire a concise model of the human body automatically. The solution to the above problems requires the development of an algorithm that: (1) integrates the processes of segmentation and fitting, (2) allows reliable shape description of the parts, (3) estimates the true location of the joints between the parts, (4) detects multiple joints, and (5) obviates the need for markers and special equipment.

As a first step towards such an algorithm, we have presented a physics-based approach to the shape and motion estimation of non-occluded chain-like structures in human body outlines (e.g., arms and legs) [5]. In this paper, we extend our technique to be able to fully segment outlines of moving humans. Initially, we assume that the human body consists of a single part. Using our physics-based framework, we fit a deformable model to the given time-varying data and we monitor the relevant model parameters. Due to intra-part occlusion (partial or full) and the relative motion of the human parts, the shape of the outline dynamically changes. In particular, tree-like structures (e.g., from the motion of the arm with respect to the torso) and chain-like structures (e.g., from the motion of the upper arm with respect to the forearm) are protruding from the outline. Identifying these structures will lead to the identification of the underlying human parts. To this end, we introduce a new method for modeling shapes with large protrusions which amounts to composition of deformable models. This method allows us to represent the shape of an outline in a compact way and to hypothesize an underlying part structure. We can verify the hypothesis by monitoring the relative motion of the composed deformable models. Once the hypothesis of multiple underlying parts is verified, we are able to identify their joint location based on the estimated shape and motion parameters. Our algorithm

for part-identification, and for the estimation of shape and motion is applied iteratively over subsequent frames until all the moving parts are identified.

In this paper, we first formulate the theory of parametric composition of geometric primitives. We then present the algorithm for shape and motion estimation for parts that gradually become visible. Finally, we present selected experimental results demonstrating our algorithm.

2 Deformable models

Following the notation in [9], the 3D position $\mathbf{x}$ of a point (with material coordinates $\mathbf{u} = (u, v)$) on a deformable model at time t, with respect to an inertial frame of reference Φ is given by $\mathbf{x}(\mathbf{u}, t) = \mathbf{c}(t) + \mathbf{R}(t)\mathbf{p}(\mathbf{u}, t)$, where $\mathbf{p}(\mathbf{u}, t)$ is the position of the given point with respect to a model-centered reference frame ϕ. In addition, $\mathbf{c}(t)$ and $\mathbf{R}(t)$ denote the position of the origin and the orientation of ϕ with respect to Φ. We further express the position $\mathbf{p}$ as: $\mathbf{p} = \mathbf{s} + \mathbf{d}$, where $\mathbf{s}(\mathbf{u}, t)$ is the model's global reference shape and $\mathbf{d}(\mathbf{u}, t)$ represents a displacement function. We define the global reference shape as $\mathbf{s} = \mathbf{T}(\mathbf{e}(\mathbf{u}; a_0, a_1, \ldots); b_0, b_1, \ldots)$, where the geometric primitive $\mathbf{e}$ (defined parametrically over $\mathbf{u}$ with global shape parameters a_i) is subjected to the global deformation $\mathbf{T}$ which depends on the parameters b_i. To represent the local deformations, we employ finite element shape functions which are tensor products of one-dimensional Hermite polynomials [5]. Therefore, $\mathbf{d} = \mathbf{S}\mathbf{q}_d$, where $\mathbf{S}$ is the shape matrix whose entries are the shape functions and $\mathbf{q}_d = (\ldots, {d_i}^\top, \ldots)^\top$ is the vector of local deformation parameters.

2.1 Parametric Composition of Geometric Primitives

To represent large protrusions or concavities and their shape evolution in a compact and intuitive way, we introduce a new representation based on the parametric composition of primitives. Intuitively, using the parametric composition of primitives, we can describe compactly the shape of their union (in the case of protrusions) or intersection (in the case of concavities). Fig. 1 depicts an example of composition (union) of two superellipsoids.

In the interest of space, we will formulate the theory of composition in 2D, to represent the shape of the boundary of the union of primitives[1]. Let $\mathbf{x}_0$ and $\mathbf{x}_1$ be two 2D parametric primitives (defined by the mappings $\mathcal{C}_0{:}[v_{0_b}, v_{0_e}) \rightarrow \mathbb{R}^2$ and $\mathcal{C}_1{:}[v_{1_b}, v_{1_e}) \rightarrow \mathbb{R}^2$ respectively - where the subscripts b and e denote the beginning and end of the domain) positioned in space so that $\mathbf{x}_1$, *intersecting primitive*, intersects $\mathbf{x}_0$, *root primitive*, at points

[1] In the 2D case the material coordinate space is one-dimensional and $\mathbf{u} = v$.

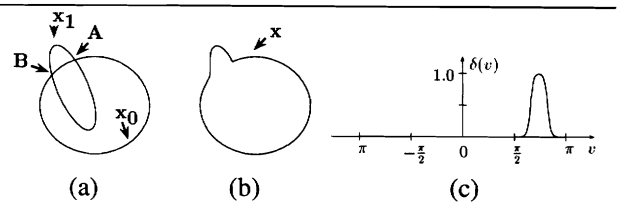


Figure 1: An example of composition of two superellipsoids: (a) depicts $\mathbf{x}_1(v_1)$ which intersects $\mathbf{x}_0(v_0)$ at points A and B, (b) depicts their composition $\mathbf{x}(v)$, and (c) depicts the composition function $\delta(v; 0.65\pi, 0.77\pi, 10)$.

A and B. The coordinates of these two points can be expressed in terms of either the material coordinate v_0 of $\mathbf{x}_0$ or the material coordinate v_1 of $\mathbf{x}_1$. Let $v_{0,\mathrm{A}}$ and $v_{0,\mathrm{B}}$ be the values of v_0, and $v_{1,\mathrm{A}}$ and $v_{1,\mathrm{B}}$ be the values of v_1 at the points A and B, respectively (Fig. 2(a,d)). Without loss of generality, we can assume that we name the points of intersection A and B so that the relation $v_{0,\mathrm{A}} < v_{0,\mathrm{B}}$ holds. Also, let I_0 be the curve segment of $\mathbf{x}_0$ which lies in the interior of the union of $\mathbf{x}_0$ and $\mathbf{x}_1$ (Fig. 2(b)), and J_0 be the curve segment of $\mathbf{x}_0$ which belongs to the boundary of their union (Fig. 2(c)). We define I_1 and J_1 in a similar way (Fig. 2(e,f)). Intuitively, in composing two primitives to represent the boundary of their union, we want to map I_0 to J_1 and J_0 to I_1. However, depending on the position of the point $\mathcal{C}_0(v_{0_b})$ (if it belongs or not to I_0), the curve I_0 can be the map of either a continuous interval or of a union of continuous intervals. For example, a superellipsoid $\mathbf{x}_0(v_0)$ is defined by the mapping $\mathcal{C}_0{:}[-\pi, \pi) \rightarrow \mathbb{R}^2$ as depicted in Fig. 2(c). When a superellipsoid $\mathbf{x}_1$ intersects $\mathbf{x}_0$, the point $\mathcal{C}_0(-\pi)$ either belongs to I_0 (Fig. 2(i,k)) or not (Fig. 2(h,j)). In the first case, I_0 is the map of the union of two continuous intervals $\mathrm{I}_0 = \{\mathbf{x}_0(v_0) : v_0 \in (v_{0,\mathrm{B}}, \pi) \cup [-\pi, v_{0,\mathrm{A}})\}$. In the second case, I_0 is the map of a single interval $\mathrm{I}_0 = \{\mathbf{x}_0(v_0) : v_0 \in (v_{0,\mathrm{A}}, v_{0,\mathrm{B}})\}$. This distinction arises from the fact that the open interval $[v_{0_b}, v_{0_e})$ and the closed curve $\mathbf{x}_0$ are not homeomorphic[2].

Based on the above, the shape $\mathbf{x}$ of the composed primitive ($\mathcal{C}{:}[v_b, v_e) \rightarrow \mathbb{R}^2$), can be defined in terms of the parameters of the defining primitives $\mathbf{x}_0$ and $\mathbf{x}_1$ as follows:

$$\mathbf{x}(v) = (1 - \delta(v))\, \mathbf{x}_0(h_0(v)) + \delta(v)\, \mathbf{x}_1(h_1(h_0(v))), \quad (1)$$

where $\delta{:}[v_b, v_e) \rightarrow [0, 1]$ is the composition function for the union of two primitives. Specifically, $\delta(v; v_{\min}, v_{\max}, c) = \sigma(c\,(v - v_{\min})) - \sigma(c\,(v - v_{\max}))$, where c is a constant that controls the shape of the function δ at the neighborhoods of $v_{\min}$ and $v_{\max}$, and $\sigma{:}\mathbb{R} \rightarrow [0, 1]$ is defined as: $\sigma(x) = \frac{1+\tanh(x)}{2}$. The function σ approximates the step function. The piecewise linear function $h_0{:}[v_b, v_e) \rightarrow [v_{0_b}, v_{0_e})$ maps the material coordinate v to v_0 and is defined as: $h_0(x) = f_1(x)$ (see appendix for the

[2] A set $\mathbf{S}$ is topologically equivalent or homeomorphic to a set $\mathbf{T}$ iff there is a 1-1 bicontinuous mapping f of $\mathbf{S}$ onto $\mathbf{T}$.

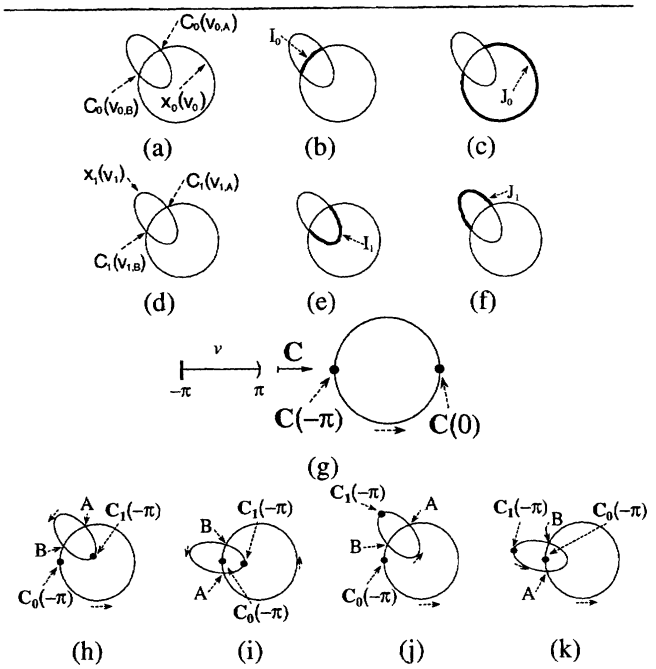


Figure 2: Figs.(a-c) illustrate the notation pertaining to $\mathbf{x}_0$, Figs.(d-f) illustrate the notation pertaining to $\mathbf{x}_1$, Fig.(g) demonstrates that the interval $[-\pi, \pi)$ and a closed curve are not topologically equivalent, and Figs.(h-k) depict possible positions of the points $\mathcal{C}_0(-\pi)$ and $\mathcal{C}_1(-\pi)$.

definition of the functions $f_i(x), i = 1 \ldots 4$). The piecewise linear function $h_1(x)$:$[v_{0_b}, v_{0_e}) \rightarrow [v_{1_b}, v_{1_e})$ maps the material coordinate v_0 to v_1. In effect, $h_1(x)$ maps I_0 to J_1 and J_0 to I_1 in order to form the union. For the definition of $h_1(x)$, four cases have to be distinguished. These four cases reflect if the curve segments I_0, J_0, I_1 and J_1 (their definition for each case are given at the appendix) are maps of a single continuous interval or of a union of two continuous intervals.

Case I: If $\mathcal{C}_0(v_{0_b}) \in \mathrm{J}_0$ and $\mathcal{C}_1(v_{1_b}) \in \mathrm{I}_1$ (Fig. 2(h)), then

$$h_1(x) = \begin{cases} f_1(x) & x \in (v_{0,\mathrm{A}}, v_{0,\mathrm{B}}) \\ f_4(x) & x \in [v_{0,\mathrm{B}}, \pi) \cup [-\pi, v_{0,\mathrm{A}}]. \end{cases}$$

The function f_1:$(v_{0,\mathrm{A}}, v_{0,\mathrm{B}}) \rightarrow (v_{1,\mathrm{A}}, v_{1,\mathrm{B}})$ maps I_0 to J_1, and the function f_4:$[v_{0,\mathrm{B}}, \pi) \cup [-\pi, v_{0,\mathrm{A}}] \rightarrow [v_{1,\mathrm{B}}, \pi) \cup [-\pi, v_{1,\mathrm{A}}]$ maps J_0 to I_1.

Case II: If $\mathcal{C}_0(v_{0_b}) \in \mathrm{I}_0$ and $\mathcal{C}_1(v_{1_b}) \in \mathrm{I}_1$ (Fig. 2(i)), then

$$h_1(x) = \begin{cases} f_2(x) & x \in (v_{0,\mathrm{B}}, \pi) \cup [-\pi, v_{0,\mathrm{A}}) \\ f_3(x) & x \in [v_{0,\mathrm{A}}, v_{0,\mathrm{B}}]. \end{cases}$$

The function f_2:$(v_{0,\mathrm{B}}, \pi) \cup [-\pi, v_{0,\mathrm{A}}) \rightarrow (v_{1,\mathrm{B}}, v_{1,\mathrm{A}})$ maps I_0 to J_1, and the function f_3:$[v_{0,\mathrm{A}}, v_{0,\mathrm{B}}] \rightarrow [v_{1,\mathrm{A}}, \pi) \cup [-\pi, v_{1,\mathrm{B}}]$ maps J_0 to I_1.

Case III: If $\mathcal{C}_0(v_{0_b}) \in \mathrm{J}_0$ and $\mathcal{C}_1(v_{1_b}) \in \mathrm{J}_1$ (Fig. 2(j)), then

$$h_1(x) = \begin{cases} f_3(x) & x \in (v_{0,\mathrm{A}}, v_{0,\mathrm{B}}) \\ f_2(x) & x \in [v_{0,\mathrm{B}}, \pi) \cup [-\pi, v_{0,\mathrm{A}}]. \end{cases}$$

The function f_3:$(v_{0,\mathrm{A}}, v_{0,\mathrm{B}}) \rightarrow (v_{1,\mathrm{A}}, \pi) \cup [-\pi, v_{1,\mathrm{B}})$ maps I_0 to J_1, and the function f_2:$[v_{0,\mathrm{B}}, \pi) \cup [-\pi, v_{0,\mathrm{A}}] \rightarrow [v_{1,\mathrm{B}}, v_{1,\mathrm{A}}]$ maps J_0 to I_1.

Case IV: If $\mathcal{C}_0(v_{0_b}) \in \mathrm{I}_0$ and $\mathcal{C}_1(v_{1_b}) \in \mathrm{J}_1$ (Fig. 2(k)), then

$$h_1(x) = \begin{cases} f_4(x) & x \in (v_{0,\mathrm{B}}, \pi) \cup [-\pi, v_{0,\mathrm{A}}) \\ f_1(x) & x \in [v_{0,\mathrm{A}}, v_{0,\mathrm{B}}]. \end{cases}$$

The function f_4:$(v_{0,\mathrm{B}}, \pi) \cup [-\pi, v_{0,\mathrm{A}}) \rightarrow (v_{1,\mathrm{B}}, \pi) \cup [-\pi, v_{1,\mathrm{A}})$ maps I_0 to J_1, and the function f_1:$[v_{0,\mathrm{A}}, v_{0,\mathrm{B}}] \rightarrow [v_{1,\mathrm{A}}, v_{1,\mathrm{B}}]$ maps J_0 to I_1.

The equations above generalize easily to the case of multiple intersecting primitives. In the following section, we will examine how the proposed representation can be used for the analysis of human outlines.

3 Active body part extraction: shape and motion estimation

The goal of our approach is to automatically segment body outlines of moving humans without using a prior model of the human body and the shape of its parts. To accomplish this goal, we request that the actor performs a set of motions according to a protocol that reveals the structure of the human body. All the motions start and end in the position where the body is erect and the arms are placed straight against the sides of the body facing medially (*body reference position*). The set of motions required to identify the shape and connectivity of the head and the limbs of an actor are described below.

Protocol of Motions

1. **Head Motion:** The actor tilts the head forward as far as possible.
2. **Left (or right) upper body extremities motions:** In the first phase of the motion, the actor lifts the left (or right) arm to the front until the arm reaches the horizontal position. Continuing from this position to the second phase, the actor rotates the arm so that the hand is facing downwards and flexes the wrist. Then, the actor bends the elbow, bringing the forearm to the vertical position.
3. **Right (or left) upper body extremities motions:** The actor lifts the left (or right) arm backwards to a comfortable position, in which the arm is not fully occluded by the torso. Then, the actor performs the left upper body extremities motions using the right arm.
4. **Left (or right) lower body extremities motions:** This motion consists of two phases. In the first phase, the actor extends the left (or right) leg to the front. When the leg reaches the maximum comfortable position (in which the

two legs do not fully occlude each other), the actor flexes the foot. Then, the actor bends the knee. In the second phase, from the position where the left (or right) leg is extended, the actor steps forward and raises the rear leg. Again, when the rear leg reaches a comfortable position, the actor flexes the foot and then bends the knee.

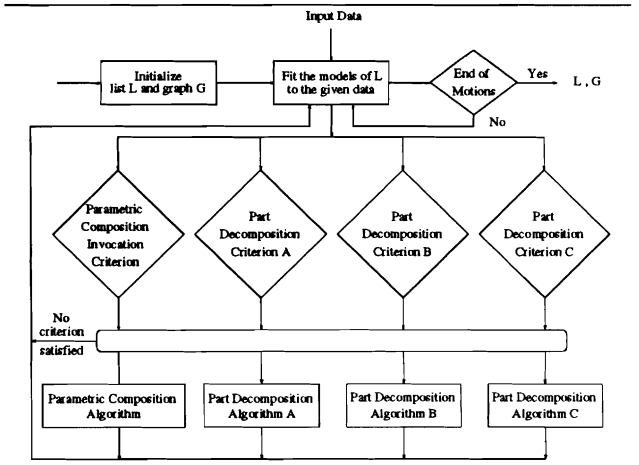


Figure 3: Flow diagram of the Human Body Part Decomposition Algorithm.

To accomplish the goal of segmentation of human body outlines, we apply the following strategy (Fig. 3).

Human Body Part Decomposition Strategy (HBPDS)

Step 1: Assume that the human body consists of a single part. Initiate a list of deformable models L (with one entry initially) that will be used to model the parts of the human body. In addition, initiate a graph G with one node. The nodes of the graph G denote the parts of the human body recovered by the algorithm. The edges of the graph denote which parts are connected by joints (Fig. 4 (a-f)).

Step 2: If not all the frames of the motion sequence have been processed, fit the models of the list L to the given data using our physics-based shape and motion estimation framework. Otherwise, output L and G.

Step 3: Perform the following steps in parallel:

a: For each model, determine if the *Parametric Composition Invocation Criterion* is satisfied.

b: For each composed model, determine if the *Part Decomposition Criterion A* is satisfied.

c: For each model, determine if the *Part Decomposition Criterion B* is satisfied.

d: For each model, determine if the *Part Decomposition Criterion C* is satisfied.

Step 4: At most one of the criteria specified in step 3 will be satisfied. Depending on the outcome, we determine which one of the following algorithms to apply.

- If 3a is satisfied, invoke the *Parametric Composition Algorithm* (as explained below). Then, go to step 2.
- If 3b is satisfied, invoke the *Part Decomposition Algorithm A*. Then, go to step 2.

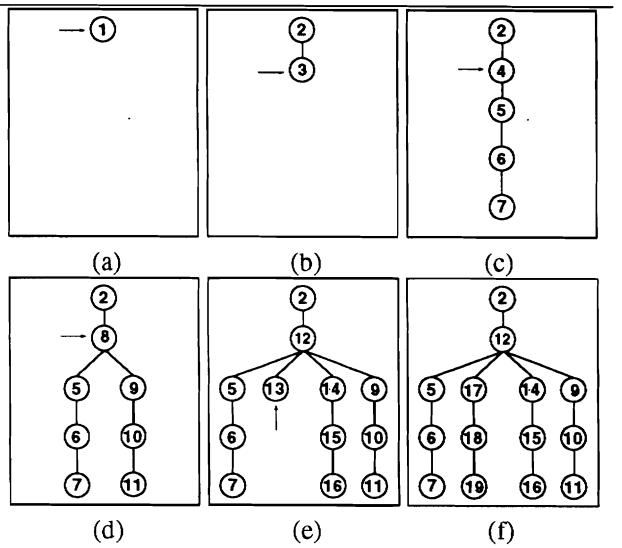


Figure 4: (a) Initially the graph of parts consists of one node. We denote by an arrow the node that has been refined in the next iteration. (b) At the completion of the head motion the graph consists of one node for the head and one for the rest of the body. (c) At the end of the motion of the left arm, the node for the rest of the body will be refined to consist of a node for the left arm, one for the left forearm, one for the left upper arm and one for the rest of the body. (d) Similarly, at the end of the motion of the right hand, our graph will contain nodes for the right upper body extremities. (e) At the end of the motion of the left leg, our graph will contain nodes for the lower body extremities. (f) The nodes of the graph at the end of all pre-specified motions.

- If 3c is satisfied, invoke the *Part DecompositionAlgorithm B*. Then, go to step 2.
- If 3d is satisfied, invoke the *Part Decomposition Algorithm C*. Then, go to step 2.
- If none of the criteria detailed in step 3 apply, go to step 2.

The Part Decomposition Criterion C and the related algorithm apply to moving chain-like structures with non occluded parts (e.g., non occluded arms and legs) and have been described at [5]. In this paper, we extend that work in order to be able to fully segment human outlines. In the following, we present the criteria and the related algorithms in more detail.

HBPDS - Step 3a: As the actor attains new postures the outline changes dynamically. Large protrusions emerge at the outline as the result of the motion of the limbs. If there is no hole present at the outline [3], we represent the protrusions as the result of composition of two primitives when the following criterion holds.

Parametric Composition Invocation Criterion: Signal the need for parametric composition of primitives if no hole is present at the outline and $\|\mathbf{d}(v,t) - \mathbf{d}(v,t_{\text{init}})\| > k$, where t_{init} is the time

[3] We will discuss the significance of a hole within the outline, when we talk about step 3c.

that the model under consideration ($\mathbf{m}_0$) was instantiated, k is an a priori defined constant and $\mathbf{d}$ is the vector of local deformations.

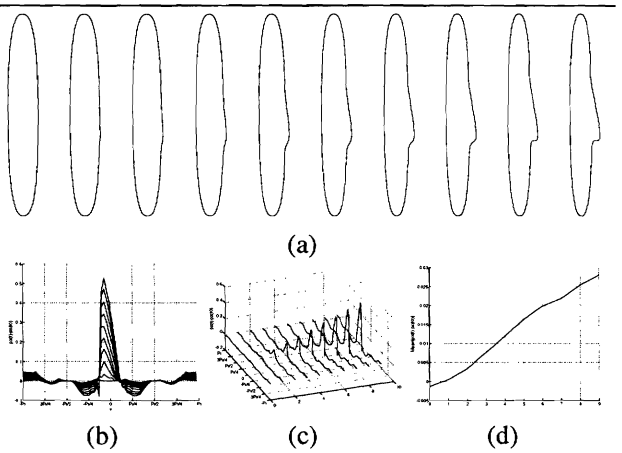


Figure 5: As the protrusions to a shape are evolving over time (Fig.(a)), we monitor the local deformations of the model (shown superimposed in Fig.(b) and over time in Fig.(c)) and the variation of the mean of their magnitude (Fig. (d)).

Parametric Composition Algorithm

I: Determine the extent of the area of the maximum variation of local deformations at $\mathbf{m}_0$.

II: Perform an eigenvector analysis on the data points that correspond to this area of the model, to approximate the parameters of a deformable model $\mathbf{m}_1$ which can describe these data.

III: Construct a composed parametric primitive $\mathbf{m}$ with defining primitives $\mathbf{m}_0$ and $\mathbf{m}_1$ (as explained in section 2.1).

IV: Update the list L, by replacing the model $\mathbf{m}_0$ with the composed model $\mathbf{m}$.

HBPDS - Step 3b: We continuously fit the models of the updated list L to the time-varying data. In the case of a composed model, we monitor the parameters of its defining primitives. If these parameters indicate that the defining primitives are moving with respect to one another then this indicates that we have two distinct parts. We determine the existence of two parts using the following criterion.

Part Decomposition Criterion A: Decompose a composed model $\mathbf{m}$ (with underlying primitives $\mathbf{m}_0$ and $\mathbf{m}_1$), when its generalized coordinates satisfy the relation $\|\Delta\theta(t) - \Delta\theta(t_{\text{init}})\| > k_A \vee \|\Delta\mathbf{q}_c(t) - \Delta\mathbf{q}_c(t_{\text{init}})\| > k_B$, where k_A and k_B are two *a priori* defined constants and t_{init} is the time that the composed model was instantiated. In addition, $\Delta\theta(t)$ is the angle between the reference frames of the defining primitives and $\Delta\mathbf{q}_c(t) = \mathbf{q}_{c_1}(t) - \mathbf{q}_{c_0}(t)$.

Part Decomposition Algorithm A

I: Construct two new models $\mathbf{n}_A$ and $\mathbf{n}_B$ using the parameters of the defining models of the composed model $\mathbf{m}$.

II: Update G and L, by replacing $\mathbf{m}$ with $\mathbf{n}_A$ and $\mathbf{n}_B$.

Later, joints between the parts are estimated employing the algorithm for joint estimation described in [5].

HBPDS - Step 3c: We now discuss the significance of the appearance of a hole within the outline. The visual event of a hole appearing within the outline indicates that parts which were initially occluded are gradually becoming visible. During this process though, the regions of the part that become visible are not contiguous, thus the appearance of the hole. Therefore, when a hole is present within the outline of a shape with large local deformations (e.g., $\mathbf{m}_0$), we do not invoke the parametric composition algorithm. We monitor the evolution of the hole, and when the hole ceases to exist (and the shape of the model exhibits large local deformations), then we invoke the Part Decomposition Algorithm C. To this end, we hypothesize that the data points that belong to the area of $\mathbf{m}_0$ with large local deformations, belong to a second part. We initialize a new deformable model $\mathbf{m}_1$, and replace $\mathbf{m}_0$ with the models $\mathbf{m}(t_{\text{init}})$ and $\mathbf{m}_1$. We fit these models to the given data, using our weighted-force assignment algorithm [5] and the physics-based framework introduced in [9]. In the following, we describe the modifications to the framework which are needed to accommodate fitting of composed models to time-varying data.

3.1 Dynamics of fitting a composed model

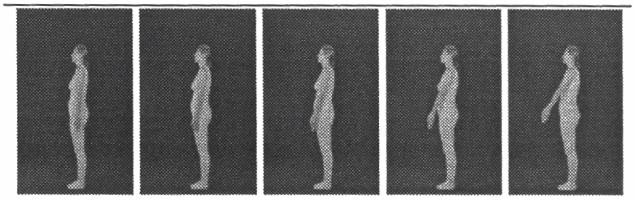
In the physics-based framework, the geometric degrees of freedom of a shape (translation, rotation, global and local parameters) form the generalized coordinates of the model. For a composed model, $\mathbf{q}_0 = (\mathbf{q}_{c_0}^\top, \mathbf{q}_{\theta_0}^\top, \mathbf{q}_{s_0}^\top, \mathbf{q}_{d_0}^\top)^\top$ and $\mathbf{q}_1 = (\mathbf{q}_{c_1}^\top, \mathbf{q}_{\theta_1}^\top, \mathbf{q}_{s_1}^\top, \mathbf{q}_{d_1}^\top)^\top$ are the generalized coordinates of the root and intersecting primitives, respectively. $\mathbf{q}_{c_i} = \mathbf{c}_i(t)$, $\mathbf{q}_{\theta_i}$ is the quaternion that represents $\mathbf{R}_i(t)$, $\mathbf{q}_{s_i}$ is the vector of global parameters, and $\mathbf{q}_{d_i}$ is the vector that specifies local deformations (i=0,1). The generalized coordinates of the composed primitive are: $\mathbf{q} = (\mathbf{q}_0^\top, \mathbf{q}_1^\top, \mathbf{q}_{\text{com}}^\top)^\top$, where $\mathbf{q}_{\text{com}} = (v_{\min}, v_{\max}, c)^\top$ are the generalized coordinates of the composition function. For the purposes of shape and motion estimation, we recover the generalized coordinates $\mathbf{q}$ in a physics-based way. Following the notation in [9], $\dot{\mathbf{x}}_0 = \mathbf{L}_0\dot{\mathbf{q}}_0$ and $\dot{\mathbf{x}}_1 = \mathbf{L}_1\dot{\mathbf{q}}_1$. From (1) we obtain:

$$\dot{\mathbf{x}} = [(1 - \delta(v))\mathbf{L}_0 \mid \delta(v)\mathbf{L}_1 \mid (\mathbf{x}_1(v) - \mathbf{x}_0(v))\frac{\partial\delta(v)}{\partial\mathbf{q}_{\text{com}}}]\,\dot{\mathbf{q}}$$

Therefore, the jacobian of the composed model depends on the jacobians of the defining models. The degree of dependence is regulated by the values of the composition function.

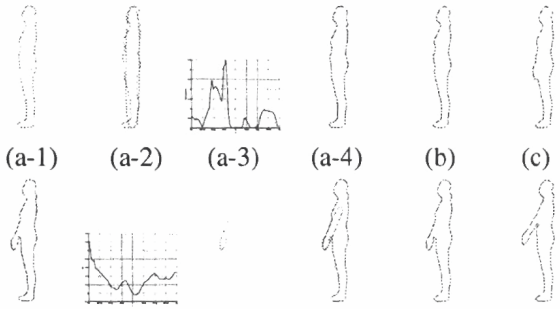
4 Experiments

We have performed several experiments demonstrating our integrated approach to segmentation, shape and non-rigid motion estimation of human body outlines.The input



(a) (b) (c) (d) (e)

Figure 6: Segmentation, shape and motion estimation of a human arm. A sample of the image sequence.

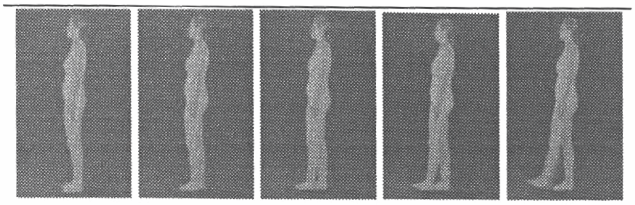


(a-1) (a-2) (a-3) (a-4) (b) (c)

(d-1) (d-2) (d-3) (d-4) (d-5) (e)

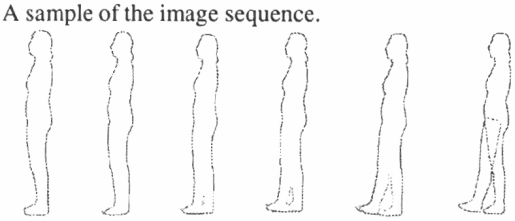
Figure 7: Segmentation, shape and motion estimation of a human arm.

to our algorithm is a sequence of monocular images from a moving human. In the current work, we restrict ourselves to the following conditions: (1) the individuals are moving in the front-parallel plane of the camera, (2) the individuals wear tight fitting clothes, and (3) the individuals are moving against a stationary background. The first restriction can be overcomed by introducing more sensors (either a pair of stereo cameras or cameras that are organized in a orthogonal configuration in space) which will allow model acquisition and tracking in 3D space. The second condition is necessary in order to acquire the true shape and the parts of the human body. To remove the third constraint, we plan to integrate into our algorithm other visual cues such as color. In this paper, we present the results from observing the actor perform the first phases of motions 2 and 4 from the protocol of motions. The outlines have been obtained by applying a Canny edge detector to the input image sequence. Figs. 6 show six frames from the image sequence. Figs. 7(a-1,a-2) show the initialization of a deformable model, and the fitted model to the first frame using only global deformations. Fig. 7(a-3) depicts the error of fit of the model depicted in Fig. 7(a-2), with respect to the material coordinate v of the geometric primitive. To reliably fit the shape using local deformations, in the areas where the error of fit exceeds a pre-specified threshold, we apply our algorithm for adaptive assignment of data points for large local deformations in the case of static images. Notice that we do not attempt to describe the shape as the result of composition of two primitives since there is no prior motion. Fig. 7(a-4) shows the finally fitted model to the first frame using global and local deformations. Figs. 7(b,c,d-1) show the results of fitting the frames in Figs. 6(b,c,d), respectively. If we plot the profile of the local deformations d_x of the model depicted in Fig. 7(d-1), we obtain Fig. 7(d-2). Fig. 7(d-3) depicts the data that we identify as the projection from another primitive. Fig. 7(d-4) depicts the recovered defining primitives of the composed model. The composed model, depicted in Fig. 7(d-5) is fitted to the data (Fig. 6(d)). Fig. 7(e) shows the fitting of the composed model to the data in Fig. 6(e). In subsequent frames, the Part Decomposition Criterion A is satisfied and we apply the Part Decomposition Algorithm A to recover the underlying parts. Later, joints between the parts are estimated employing the algorithm for joint estimation described in [5].



(a) (b) (c) (d) (e)

Figure 8: Segmentation, shape and motion estimation of a human leg. A sample of the image sequence.



(a) (b) (c) (d) (e) (f)

Figure 9: Segmentation, shape and motion estimation of a human leg.

Figs. 8 show five frames from the image sequence, in which the actor moves her leg. Figs. 9(a-d) depict the models fitted to the data in Figs. 8(a-d). The Part Decomposition Criterion C is satisfied for the the data in Fig. 8(e) so we invoke the Part Decomposition Algorithm C. The models recovered are depicted in Fig. 8(f).

5 Conclusion

We have presented a novel, integrated approach to identifying the parts of a human body and to estimating their shape and motion. Initially, we assumed that the outline is the result of the projection of a single part and we fitted a deformable model to the given data using our physics-based framework. As the actor moved and attained new postures, large protrusions emerged on the outline. To capture the complex shape of the outline, we employed a new representation consisting of composition of deformable models. By monitoring the evolution of shape and motion parameters of

the composed models, the joint locations of the underlying human parts were identified. Our physics-based algorithm couples the processes of segmentation, shape and motion estimation. This coupling produced very encouraging results on the decomposition and the estimation of shape of motion of the human parts without the need for a prior model or markers.

Appendix

In this section, first we will present the values of the I_0, J_0, I_1 and J_1 for the four different cases of composition. Then, we will define the linear functions that allow mappings between intervals.

The values of the I_0, J_0, I_1 and J_1 for the four different cases of composition are the following:

Case I: $I_0 = \{\mathcal{C}_0(v_0) : v_0 \in (v_{0,A}, v_{0,B})\}$,
$J_0 = \{\mathcal{C}_0(v_0) : v_0 \in [v_{0,B}, \pi) \cup [-\pi, v_{0,A}]\}$,
$I_1 = \{\mathcal{C}_1(v_1) : v_1 \in [v_{1,B}, \pi) \cup [-\pi, v_{1,A}]\}$, and
$J_1 = \{\mathcal{C}_1(v_1) : v_1 \in (v_{1,A}, v_{1,B})\}$.

Case II: $I_0 = \{\mathcal{C}_0(v_0) : v_0 \in (v_{0,B}, \pi) \cup [-\pi, v_{0,A})\}$,
$J_0 = \{\mathcal{C}_0(v_0) : v_0 \in [v_{0,A}, v_{0,B}]\}$,
$I_1 = \{\mathcal{C}_1(v_1) : v_1 \in [v_{1,A}, \pi) \cup [-\pi, v_{1,B}]\}$, and
$J_1 = \{\mathcal{C}_1(v_1) : v_1 \in (v_{1,B}, v_{1,A})\}$.

Case III: $I_1 = \{\mathcal{C}_1(v_1) : v_1 \in [v_{1,B}, v_{1,A}]\}$,
$J_1 = \{\mathcal{C}_1(v_1) : v_1 \in (v_{1,A}, \pi) \cup [-\pi, v_{1,B})\}$, I_0 and J_0 are the same as case I.

Case IV: $I_1 = \{\mathcal{C}_1(v_1) : v_1 \in [v_{1,A}, v_{1,B}]\}$,
$J_1 = \{\mathcal{C}_1(v_1) : v_1 \in (v_{1,B}, \pi) \cup [-\pi, v_{1,A})\}$, I_0 and J_0 are the same as case II.

In the following, we will define linear functions that allow mappings between intervals. Let $A = [a_b, a_e)$, $B = [b_b, b_e)$, $C = [c_b, c_e)$ and $D = [d_b, d_e)$ be four continuous intervals with corresponding lengths $l_A = (a_e - a_b), l_B = (b_e - b_b), l_C = (c_e - c_b)$ and $l_D = (d_e - d_b)$.

• To linearly map A to C, we we define the function f_1:$A \rightarrow C$ such that:

$$f_1(x) = \left(\frac{l_C}{l_A}\right) x + \frac{a_e c_b - a_b c_e}{l_A}$$

• To linearly map the union $A \cup B$ (assuming that $a_e \neq b_b$) to C, we define the function f_2:$A \cup B \rightarrow C$ such that:

$$f_2(x) = \begin{cases} c_b + \lambda\,(x - a_b) & x \in A \\ c_b + \lambda\, l_A + \lambda\,(x - b_b) & x \in B \end{cases}$$

where $\lambda = \frac{l_C}{(l_A + l_B)}$.

• To linearly map A to the union $B \cup C$ (where $b_e \neq c_b$), we define the function f_3:$A \rightarrow B \cup C$ such that:

$$f_3(x) = \begin{cases} b_b + p(x) & p(x) < l_B \\ c_b - l_B + p(x) & p(x) \geq l_B \end{cases}$$

where $p(x) = \lambda(x - a_b)$ and $\lambda = \frac{(l_B + l_C)}{l_A}$.

• To linearly map the union $A \cup B$ to the union $C \cup D$, we define the function f_4:$A \cup B \rightarrow C \cup D$ such that:

$$f_4(x) = \begin{cases} c_b + p(x) & p(x) < l_C \\ d_b - l_C + p(x) & p(x) \geq l_C \end{cases}$$

where the function $p(x)$ is defined as:

$$p(x) = \begin{cases} \lambda\,(x - a_b) & x \in A \\ \lambda\,(x + l_A - b_b) & x \in B \end{cases}$$

and $\lambda = \frac{(l_C + l_D)}{(l_A + l_B)}$.

Acknowledgments: This work was supported in part by the following grants: ARPA Grant N00014-92-J-1647, ARO Grant DAAL03-89-C-0031PRI, NSF Grants CISE/CDA-88-22719, IRI-9309917;The Gerondelis Foundation and The Bodosakis Foundation.

References

[1] K. Akita. Image sequence analysis of real world human motion. *Pattern Recognition*, 17:73–83, 1984.

[2] C. I. Attwood, G. D. Sullivan, and K. D. Baker. Model based recognition of human posture using single synthetic images. In *Proceedings of the Fifth Alvey Vision Conference*, pp. 25–30, Reading,UK, 1989. University of Reading.

[3] D. D. Hoffman and B. E. Flinchbaugh. Interpretation of biological motion. *Biological Cybernetics*, 42:195–204, 1982.

[4] D. Hogg. Model-based vision: a program to see a walking person. *Image and Vision Computing*, 1(1):5–20, 1983.

[5] I. A. Kakadiaris, D. Metaxas, and R. Bajcsy. Active part-decomposition, shape and motion estimation of articulated objects: A physics-based approach. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 980–984, Seattle,WA, June 21-23 1994.

[6] S. Kurakake and R. Nevatia. Description and tracking of moving articulated objects. In *Proc. of the IEEE International Conference on Pattern Recognition*, pp. 491–495, 1992.

[7] M. K. Leung and Y. H. Hang. Human body motion segmentation in a complex scene. *Pattern Recognition*, 20(1):55–64, 1987.

[8] M. K. Leung and Y. H. Hang. A region based approach for human body motion analysis. *Pattern Recognition*, 20(3):321–339, 1987.

[9] D. Metaxas and D. Terzopoulos. Shape and nonrigid motion estimation through physics-based synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6):580–591, June 1993.

[10] J. O'Rourke and N. I. Badler. Model-based image analysis of human motion using constraint propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(6):522–536, 1980.

[11] A. Pentland. Automatic extraction of deformable part models. *International J. on Computer Vision*, 4:107–126, 1990.

[12] R. J. Quian and T. S. Huang. Motion analysis of articulated objects. In *Image Understanding Workshop*, pp. 549–553, San Diego, CA, January 1992.

[13] R. F. Rashid. Toward a system for the interpretation of moving light displays. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(6):574–581, 1980.

[14] J. M. Rehg and T. Kanade. Visual tracking of high DOF articulated structures: an application to human hand tracking. In Jan-Olof Eklundth, editor, *European Conference on Computer Vision*, pp. 35–46, Stockholm, Sweden, May 2–6 1994.

[15] K. Rohr. Towards model-based recognition of human movements in image sequences. *Computer Vision, Graphics, and Image Processing: Image Understanding*, 59(1):94–115, January 1994.

[16] F. Solina and R. Bajcsy. Recovery of parametric models from range images: The case for superquadrics with global deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(2):131–146, 1990.

[17] V. T. Tsukiyama and Y. Shirai. Detection of the movements of persons from a sparse sequence of TV images. *Pattern Recognition*, 18:207–213, 1985.

[18] J. A. Webb and J. K. Aggarwal. Structure from motion of rigid and joined objects. In *Proc. of the International Joint Conf. on Artificial Intelligence*, pp. 686–691, 1981.

[19] M. Yamamoto and K. Koshikawa. Human motion analysis based on a robot arm model. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 664–665, Hawaii, 3 - 6 June 1991.