11-1-1994

Active Part-Decomposition, Shape and Motion Estimation of Articulated Objects: A Physics-based Approach

Ioannis A. Kakadiaris
University of Pennsylvania

Dimitris Metaxas
University of Pennsylvania, dnm@central.cis.upenn.edu

Ruzena Bajcsy
University of Pennsylvania

The Institute For Research In Cognitive Science

Active Part-Decomposition, Shape and Motion Estimation of Articulated Objects: A Physics-based Approach

by

Ioannis A. Kakadiaris
Dimitri Metaxas
Ruzena Bajcsy

University of Pennsylvania
3401 Walnut Street, Suite 400C
Philadelphia, PA 19104-6228

November 1994

P

E

N

N

University of Pennsylvania
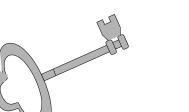Founded by Benjamin Franklin in 1740

IRCS Report 94-18

Active Part-Decomposition, Shape and Motion Estimation of Articulated Objects: A Physics-based Approach

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 a novel, robust, integrated approach to segmentation shape and motion estimation of articulated objects. Initially, we assume the object consists of a single part, and we fit a deformable model to the given data using our physics-based framework. As the object attains new postures, we decide based on certain criteria if and when to replace the initial model with two new models. These criteria are based on the model's state and the given data. We then fit the models to the data using a novel algorithm for assigning forces from the data to the two models, which allows partial overlap between them and determination of joint location . This approach is applied iteratively until all the object's moving parts are identified. Furthermore, we define new global deformations and we demonstrate our technique in a series of experiments, where Kalman filtering is employed to account for noise and occlusion.

1   Introduction

The systematic identification of an articulated object's parts has been a longstanding research topic in computer vision. Accomplishing this task using a single image is an underconstrained problem. For example, when we observe a human arm in a posture as in Figs. 1(a-b), assuming no prior knowledge about its structure, we cannot decide whether it is composed of multiple parts. Similarly, based on Fig. 1(c), we may conclude that it is a bent object. In this paper, we develop a new technique to reliably identify an articulated object's parts, joints, shape and motion. Our technique uses an object's motion sequence to provide the necessary constraints for the above tasks.

Despite the large body of work on segmentation, shape and motion estimation of articulated objects, most existing techniques either assume a priori knowledge of an object's

parts [7, 11, 4] or determine its parts under certain assumptions [12, 2, 6, 3, 8, 5]. In all of the above techniques, the process of segmentation and the process of shape and motion estimation are decoupled leading to possible lack of robustness and inaccuracies in shape and motion estimation.
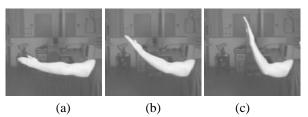

(a) (b) (c)
Figure 1: Different postures of a moving human arm

In this paper, we present our integrated approach to segmentation, shape and motion estimation of complex articulated objects whose parts form open chains. Our physics-based algorithm couples the processes of segmentation, shape and motion estimation. This coupling allows the robust extraction of parts and the estimation of object shape and motion. Our input is a sequence of monocular images from a moving articulated object. We initially assume that the data in the first frame belong to a single-part object and we fit a single deformable model. As the model deforms to fit data from subsequent frames we decide when to initially split the model to two new models based on certain criteria. These criteria depend on the model's state and the given data. In order to achieve partial overlap between the above two models at the end of the fitting process, we devise a new algorithm for assigning weighted forces from a given datapoint to each of the two models. These weights are computed based on the theory of fuzzy clustering and allow partial overlap between parts. Our algorithm for part-identification, shape and motion estimation is applied iteratively over subsequent frames until all the object's moving parts are identified. In order to cope

with occlusion, we incorporate a Kalman filter to our dynamic fitting algorithm to predict the location of the data at the next frame.

2 Deformable models

Recently, we developed [4] a physics-based framework which provides deformable models and robust techniques for inferring shape and motion from noisy data. Following the notation in [4] the position $\mathbf{x}$ of a point on the deformable model is given by $\mathbf{x} = \mathbf{c} + \mathbf{R}\mathbf{p}$ where $\mathbf{c}$ and $\mathbf{R}$ are the translation and rotation of the model frame with respect to the reference frame, while $\mathbf{p}$ is the position of the given point with respect to the model frame. Furthermore, $\mathbf{p} = \mathbf{s} + \mathbf{d}$, where $\mathbf{d}$ represents local deformations and $\mathbf{s}$ is the model's global reference shape. This global reference shape is defined as $\mathbf{s} = \mathbf{T}(\mathbf{e}(\mathbf{u}; a_0, a_1, \ldots); b_0, b_1, \ldots)$, where the geometric primitive $\mathbf{e}$ is subjected to the global deformation $\mathbf{T}$ which depends on the parameters b_i. We extend the class of allowable global deformations to include a parameterized *piecewise bending deformation* that ensures constant curvature along the major axis of bending. This definition is inspired from [1] and is useful for many natural and man-made objects.

The domain of bending is a bounded subspace of the Euclidean space R^3. This domain is partitioned into three non-intersecting zones: the *fixed zone*, the *bending zone* and the *relocation zone*. The fixed zone remains unchanged during bending. In the bending zone, parameter b_0 denotes the radius of curvature. The range of the bending zone is controlled by parameters b_1 and b_2. The center of the bend occurs at $(b_1 + b_2)/2$. The relocation zone is translated and rotated rigidly. The bending angle θ is constant at the extremities and changes linearly in the bending zone. Specifically:

$$\theta = b_0 \left(\hat{e}_2 - \frac{(b_1 + b_2)}{2} \right),$$

$$\hat{e}_2 = \begin{cases} b_1, & e_2 \leq b_1 \\ e_2, & b_1 < e_2 < b_2 \\ b_2, & e_2 \geq b_2 \end{cases}. \quad (1)$$

The isotropic bending deformation $\mathbf{s} = \mathbf{T}_b(\mathbf{e}; b_0, b_1, b_2)$ along a centerline parallel to the x-axis of a primitive $\mathbf{e} = (e_1, e_2, e_3)^T$ is given by:

$$s_1 = cos(\theta)(e_1 - \frac{1}{b_0}) + \frac{1}{b_0} + sin(\theta)(e_2 - \hat{e}_2)$$

$$s_2 = -sin(\theta)(e_1 - \frac{1}{b_0}) + \frac{b_1 + b_2}{2} + cos(\theta)(e_2 - \hat{e}_2)$$

$$s_3 = e_3. \quad (2)$$

Our definition, versus the one presented in [1] allows us to decouple the recovery of the rotation and bending parameters during model fitting. To approximate the thin plate under tension deformation energy [4], suitable for C^1 continuous model surface, we employ the finite element method to discretize our deformable models into a set of connected element domains. Our implementation is based on the use of a new class of shape functions which are tensor products of one-dimensional Hermite polynomials:

$$\begin{array}{llll} H_1^0(\xi) & = & 1 - 3\xi^2 + 2\xi^3 & H_1^1(\xi) = \xi(\xi - 1)^2 \\ H_2^0(\xi) & = & \xi^2(3 - 2\xi) & H_2^1(\xi) = \xi^2(\xi - 1), \end{array}$$

where the subscripts are related to the two endpoints of the one-dimensional segment and the superscripts, 0 and 1, denote the association of a basis function to a nodal variable and a nodal derivative, respectively. The finite element nodal degrees of freedom are the nodal displacements and their derivatives.

3 Active part-identification, shape and motion estimation

Instead of estimating the shape and motion of complex objects under the assumption of prior segmentation, our technique allows active, simultaneous segmentation and fitting. To identify the object's parts, we use a sequence of images which contain different postures of the moving object. When we observe an articulated object in a posture where the articulations are not detectable, we assume initially that the object 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. As the object moves and attains new postures, we decide if and when to replace the initial model with two new models. This decision is based on the error of fit, the rate of change and magnitude of the bending deformation and the continuity of the given data within the bending region. The first two criteria are necessary to signal that the global deformations are inadequate to represent the object's shape accurately and that there is a shape change over time. However, they are not sufficient to signal that there is more than one part. For example, if the image sequences are taken from a bending elastic object, then the error from fitting the data using only global deformations should not lead us to the conclusion that there exist two parts. The reason is that if we allow local deformations, we can minimize the error of fit. The third criterion, the detection of a discontinuity in the first derivative of the given data within the bending region, is what distinguishes an elastic object from an articulated object like a robot arm, a human arm or a human finger.

When the above criteria are met, we replace the initial model with two new models. We identify the data that correspond to the fixed, bending and relocation zones of the initial model based on the estimated bending parameters b_1 and b_2 and the image projection assumptions (orthographic or perspective). We then initialize the two models based on the data that correspond to the fixed and the relocation zones

of the initial model. However, the datapoints that correspond to the bending region of the initial model are marked as *orphan datapoints* since it is uncertain as to which of the two new models they should be assigned. This is necessary since we do not know in advance the shape of each of the two models. Our goal then is to fit the two new models to the given data. In addition, we would like them to fit in a way that allows partial overlap between the two parts. Since we know to which model the data in the fixed and the relocation zones belong, we use our previously developed algorithm for assigning forces from datapoints to points on the model. To assign forces from the orphan data to the two models, we use a novel algorithm that allows the weighted assignment of a given orphan datapoint to both deformable models. We compute these weights, whose sum is always equal to one, by minimizing an appropriately selected energy expression. Once we compute all the forces from the datapoints to the two models, we estimate the shape and motion of the two new models using our physics-based framework [4]. Our algorithm can be applied under both orthographic and perspective projections based on recent extensions presented in [5].

3.1 Weighted force-assignment based on fuzzy clustering

Since we do not know to which model each orphan datapoint should be assigned, we developed a new algorithm for assigning forces from an orphan datapoint to each of the two deformable models inspired by the theory of fuzzy clustering [9]. Clustering techniques are normally applied to feature space, but in certain cases they can be directly applied in image space. When feasible, direct application of clustering algorithms may have advantages over feature space approaches. Examples of such advantages is the applicability to sparse data so that there is no need to extract features and the lower sensitivity to noise. In the context of treating the image space itself as the feature space, the problem of assigning forces from the orphan datapoints to the models is viewed as a direct clustering problem. Most methods, which are based on objective function minimization may be classified into two categories: *hard* or *fuzzy*. In hard or crisp methods, each sample vector strictly belongs to one and only one cluster. In fuzzy methods it is shared to varying degrees among several clusters. In our algorithm, fuzziness is related to the uncertainty, by introducing a datapoint's degree of membership in a particular model. Our algorithm can be viewed as gradually decreasing the fuzziness of the associations. In this way, a datapoint exerts a force to each of the two deformable models instead of only to one. Each datapoint exerts a force to the point on the surface of each model which has minimal distance from it. The force is proportional to the distance between the datapoint and the selected point on the model. Each force
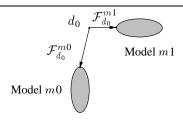

Figure 3: Force Assignment

is subsequently multiplied by a *certainty* weight. We compute these certainty weights by minimizing an appropriately defined energy term subject to the constraint that the sum of the certainty weights is one. Intuitively, the certainty weights represent the degree of membership of a datapoint to a given model. Thus our algorithm assigns a higher certainty weight to the force exerted from the datapoint to a point on a model that has the minimum distance from the datapoint. Even though we highlighted our algorithm for the case of two models, our algorithm is applicable for assigning a datapoint to any number of deformable models.

We will now formally present our algorithm for the case of a given datapoint that can be assigned to an arbitrary number of models m. We assume that we have a set of datapoints $\mathbf{d}_i, i = 1..n$, and we want to find the certainty weights for the forces that will be exerted from each datapoint $\mathbf{d}_i$, to each of the deformable models $j = 1...m$. We then denote the weight of the force from datapoint $\mathbf{d}_i$ to model j as $p_{d_{i,j}}$. Let also $\mathbf{x}_i^j$ be the point on model j that corresponds to datapoint $\mathbf{d}_i$ and let $p_{x_i^j,k}$ be the certainty measure that the model point $\mathbf{x}_i^j$ belongs to model $k = 1...m$ (its value is 1 if $k = j$ and 0 otherwise). We then define the following energy term

$$E_w = \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{m} \frac{(p_{d_{i,j}} - p_{x_i^j,k})^2}{||\mathbf{d}_i - \mathbf{x}_i^j||^2} \qquad (3)$$

that we want to minimize with respect to $p_{d_{i,j}}$ subject to the constraints $\sum_{j=1}^{m} p_{d_{i,j}} = 1, \ i = 1...n$. We perform the minimization for each datapoint $\mathbf{d}_i$ using the method of Lagrange multipliers and minimize the following formulas

$$E_{w,i} = \sum_{j=1}^{m} \sum_{k=1}^{m} \frac{(p_{d_{i,j}} - p_{x_i^j,k})^2}{||\mathbf{d}_i - \mathbf{x}_i^j||^2} - \lambda(\sum_{j=1}^{m} p_{d_{i,j}} - 1) \qquad (4)$$

with respect to $p_{d_{i,j}}$. The minimum of those formulas is computed by setting $\partial E_{w,i}/\partial p_{d_{i,j}} = 0, i = 1...n$. After some algebraic manipulation we compute the $p_{d_{i,j}}, i = 1...n$, based on the following formulas

$$p_{d_{i,j}} = \frac{1}{m} + \frac{\sum_{k=1}^{m} ||\mathbf{d}_i - \mathbf{x}_i^j||^2 (p_{x_i^j,k} - \frac{1}{m})}{\sum_{j=1}^{m} ||\mathbf{d}_i - \mathbf{x}_i^j||^2}. \qquad (5)$$

Once we compute the certainty weights $p_{d_{i,j}}(i = 1...n, j = 1...m)$, we multiply them with the corresponding

distance between $\mathbf{d}_i$ and $\mathbf{x}_i^j$ to compute the resulting forces $\mathcal{F}_{d_i}^j = p_{d_{i,j}} \ (\mathbf{d}_i - \mathbf{x}_i^j)$. An important property of our new force assignment algorithm is that it allows partial overlap between the two models at a joint. Therefore, we can determine the joint location in an articulated object using the following algorithm.

3.2 Determination of joint location

Let's assume that we have estimated the shape and motion of the two parts of an articulated object at times t and $t + \delta t$, by fitting two models m_0 and m_1. Then we want to identify the location of their common joint. Following the notation in [4], the unknown location of the center of the joint can be expressed in terms of the parameters of model m_0 at times t and $t + \delta t$ as:

$$
\begin{aligned}
\mathbf{x}_0(t) &= \mathbf{c}_0(t) + \mathbf{R}_0(t)p_0 \\
\mathbf{x}_0(t + \delta t) &= \mathbf{c}_0(t + \delta t) + \mathbf{R}_0(t + \delta t)p_0
\end{aligned}
\tag{6}
$$

and with respect to the parameters of model m_1 at times t and $t + \delta t$ as:

$$
\begin{aligned}
\mathbf{x}_1(t) &= \mathbf{c}_1(t) + \mathbf{R}_1(t)p_1 \\
\mathbf{x}_1(t + \delta t) &= \mathbf{c}_1(t + \delta t) + \mathbf{R}_1(t + \delta t)p_1
\end{aligned}
\tag{7}
$$

Under the obvious assumption that $\mathbf{x}_0(t) = \mathbf{x}_1(t)$ and $\mathbf{x}_0(t + \delta t) = \mathbf{x}_1(t + \delta t)$ and by subtracting the above two equations, we arrive at the following system of equations, with unknowns the locations $\mathbf{p}_0$ and $\mathbf{p}_1$ of the joint with respect to the model reference frames of the two models,

$$
\begin{bmatrix} \mathbf{R}_0(t) & -\mathbf{R}_1(t) \\ \mathbf{R}_0(t + \delta t) & -\mathbf{R}_1(t + \delta t) \end{bmatrix} \begin{bmatrix} \mathbf{p}_0 \\ \mathbf{p}_1 \end{bmatrix} =
$$

$$
\begin{bmatrix} \mathbf{c}_1(t) - \mathbf{c}_0(t) \\ \mathbf{c}_1(t + \delta t) - \mathbf{c}_0(t + \delta t) \end{bmatrix}
\tag{8}
$$

which is easily solved. We follow a Kalman filter based approach [4], if the location of the joint varies between frames due to noise in the data. Therefore, we can robustly estimate the locations of the joints of an articulated object.

3.3 Coping with occlusion

To cope with occlusion, we use a continuous extended Kalman filter [4] to predict the location of the data at the next time step, in addition to filtering the noise. The prediction is based on the magnitude of the estimated parameter derivatives which define and allow a spatio-temporal search space (our parameters are associated with both the shape and the motion of the model). In this way we can ignore spurious edges in both space and time that get introduced when another object temporarily occludes part of our object.
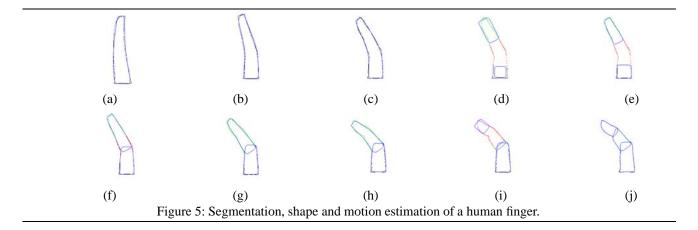

Figure 4: Segmentation, shape and motion estimation of a human finger. A sample of the image sequence.

4 Experiments

We performed experiments demonstrating our integrated approach to segmentation, shape and nonrigid motion estimation from motion image data obtained >from a robot arm, a human arm with occlusion and a human finger. Due to lack of space, we only present the results from the last two experiments.

We use image data obtained from the planar motion of a bending human finger (Fig. 4). Fig. 5(a) shows the final fitted model to the first frame using only global deformations. Fig. 5(b) shows the model fitted to a subsequent image frame. Fig. 5(c) shows the model fitted to the image frame where the partitioning criteria are satisfied and the hypothesis that the object is comprised from two parts is generated. Figs. 5(d-f) demonstrate the fitting of the two new models to the image data. Fig. 5(d) shows the initialization of the new models, Fig. 5(e) shows an intermediate step in the fitting process, while Fig. 5(f) shows the finally fitted models. The overlap between the two models allows us to compute robustly the location of the joint over several frames and place a point-to-point constraint between the two models. Fig. 5(g) shows the models fitted to a new frame, while Fig. 5(h) shows the models fitted to the frame where the partitioning criteria are satisfied for the upper model and the hypothesis that the upper model should be replaced by two new models is generated. Fig. 5(i) shows the initialization of the two new models based on our technique, while Fig. 5(j) shows all three models fitted to the given data.

Finally, we tested our algorithm using data obtained from a human arm which is occluded during its planar motion (Figs. 6(a-c)). Fig. 6(d) shows the data from an intermediate image frame, where the existence of two models has been established. The location of the joint has been determined and a point-to-point constraint enforces the contact of the two models. Fig. 6(e) shows data from a subsequent frame where partial occlusion occurs; it can be seen in the form of additional edge points. Figs. 6(f-g) show the previous position of the models and the models fitted to the new data while ignoring the additional datapoints due to occlusion through the use of the predictive power of the Kalman filter. Fig. 6(h) shows data from a subsequent frame where partial occlusion occurs resulting in missing contour points, while Fig. 6(i) shows the models
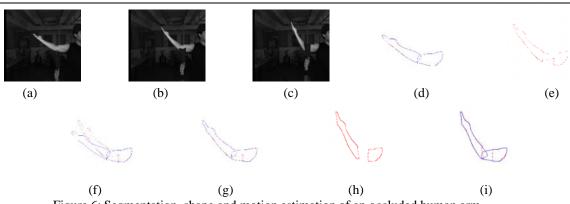
Figure 5: Segmentation, shape and motion estimation of a human finger.

Figure 6: Segmentation, shape and motion estimation of an occluded human arm.

fitted to these data.

5 Conclusion

We have presented a novel integrated approach to segmentation, shape and motion estimation. Based on certain criteria that depend on the model's state and the given image sequence, our physics-based estimation technique allows the iterative part-identification, shape and motion estimation of articulated objects whose parts form open chains. Our algorithm allows identification of joint location and can cope with occlusion. We are currently extending our algorithm to allow segmentation of more complex shapes like human bodies.

References

[1] A. Barr, "Global and Local Deformations of Solid Primitives", *Computer Graphics*, 18:21–30, 1984.

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

[3] S. Kurakake and R. Nevatia, "Description and Tracking of Moving Articulated Objects", *Proceedings of the 11th International Conference on Pattern Recognition*,pp. 491-495, 1992.

[4] D. Metaxas and D. Terzopoulos," Shape and Nonrigid Motion Estimation Through Physics-Based Synthesis" *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(6):580–591, June, 1993.

[5] D. Metaxas and S. Dickinson, "Integration of Quantitative and Qualitative Techniques for Deformable Model Fitting from Orthographic, Perspective, and Stereo Projections", it Proc. 4th International Conference on Computer Vision (ICCV'93), pp. 641-649, Berlin, Germany, May, 1993.

[6] A. Pentland, "Automatic Extraction of Deformable Part Models", *International Journal of Computer Vision*, 4:107–126, 1990.

[7] A. Pentland and B. Horowitz. Recovery of Non-rigid Motion and Structure. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(7):730–742, 1991.

[8] R. J. Quian and T. S. Huang, "Motion Analysis of Articulated Objects",*Image Understanding Workshop*,San Diego, CA., Jan.,92, pp. 549–553.

[9] E. Ruspini, "Numerical methods for Fuzzy Clustering", *Information Science*,6:273–284, 1972.

[10] D. Terzopoulos and D. Metaxas, "Dynamic 3D Models with Local and Global Deformations: Deformable Superquadrics",*IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(7):703–714, 1991.

[11] M. Yamamoto and K. Koshikawa, "Human Motion Analysis Based on A Robot Arm Model", *IEEE Computer Vision and Pattern Recognition Conference (CVPR'91)*, pp. 664–665, 1991.

[12] J. A. Webb and J. K. Aggarwal, "Structure from motion of rigid and joined objects", *Proc. International Joint Conference in Artificial Intelligence*, pp.686-691, 1981.