In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Madison, WI, June 2003.

A Bayesian Approach to Image-Based Visual Hull Reconstruction

Kristen Grauman, Gregory Shakhnarovich, Trevor Darrell

Artificial Intelligence Laboratory
Massachusetts Institute of Technology
{*kgrauman, gregory, trevor*}*@ai.mit.edu*

Abstract

We present a Bayesian approach to image-based visual hull reconstruction. The 3-D shape of an object of a known class is represented by sets of silhouette views simultaneously observed from multiple cameras. We show how the use of a class-specific prior in a visual hull reconstruction can reduce the effect of segmentation errors from the silhouette extraction process. In our representation, 3-D information is implicit in the joint observations of multiple contours from known viewpoints. We model the prior density using a probabilistic principal components analysis-based technique and estimate a maximum a posteriori reconstruction of multi-view contours. The proposed method is applied to a dataset of pedestrian images, and improvements in the approximate 3-D models under various noise conditions are shown.

1. Introduction

Reconstruction of 3-D shape using the intersection of object silhouettes from multiple views can yield a surprisingly accurate shape model, if accurate contour segmentation is available. Algorithms for computing the visual hull of an object have been developed based on the explicit geometric intersection of generalized cones [11]. More recently methods that perform resampling operations purely in the image planes have been developed [14], as well as approaches using weakly calibrated or uncalibrated views [18, 25].

Visual hull algorithms have the advantage that they can be very fast to compute and re-render, and they are also much less expensive in terms of storage requirements than volumetric approaches such as voxel carving or coloring [10, 19, 21]. With visual hulls view-dependent re-texturing can be used, provided there is accurate estimation of the alpha mask for each source view [15]. When using these techniques a relatively small number of views (4-8) is often sufficient to recover models that appear compelling and are useful for creating real-time virtual models of objects and people in the real world, or for rendering new images for view-independent recognition using existing view-dependent recognition algorithms [20].

Unfortunately most algorithms for computing visual hulls are deterministic in nature, and they do not model any uncertainty that may be present in the observed contour shape in each view. They can also be quite sensitive to segmentation errors: since the visual hull is defined as the 3-D shape which is the intersection of the observed silhouettes, a small segmentation error in even a single view can have a dramatic effect on the resulting 3-D model (see Figure 4).

Traditional visual hull algorithms (e.g., [14]) have the advantage that they are general – they can reconstruct any 3-D shape which can be projected to a set of silhouettes from calibrated views. While this is a strength, it is also a weakness of the approach. Even though parts of many objects cannot be accurately represented by a visual hull (e.g, concavities), the set of objects that can be represented is very large, and often larger than the set of objects that will be physically realizable. Structures in the world often exhibit local smoothness, which is not accounted for in deterministic visual hull algorithms [1]. Additionally, many applications may have prior knowledge about the class of objects to be reconstructed, e.g. pedestrian images as in the gait recognition system of [20]. Existing algorithms cannot exploit this knowledge when performing reconstruction or re-rendering.

In this paper we show how to formulate a probabilistic version of an image-based visual hull reconstruction, and enforce a class-specific prior shape model on the reconstruction. We learn a probability density of possible 3-D shapes, and model the observation uncertainty of the silhouettes seen in each camera. From these we compute a Bayesian estimate of the visual hull given the observed silhouettes. We use an explicit image-based algorithm, and define our prior shape model as a density over the set of object contours in each view. We restrict our focus to reconstructing a single object represented by a closed contour

[1] In practice many implementations use preprocessing stages with morphological filters to smooth segmentation masks before geometric intersection, but this may not reflect the statistics of the world and could lead to a shape bias.

in each view; this simplifies certain steps in contour processing and representation. It is well known that the probability densities of contour models for many object classes can be efficiently represented as linear manifolds [1, 2, 4], which can be computed using probabilistic principal component analysis (PPCA) techniques [22]. In essence, we extend this approach to the case of multiple simultaneous views used for visual hull reconstruction.

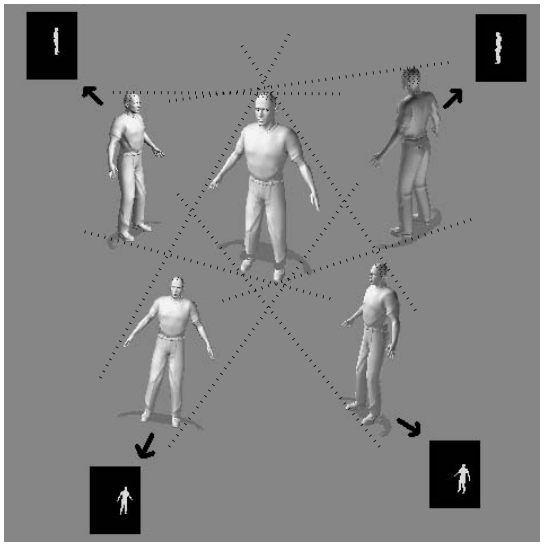


Figure 1: Schematic illustration of the geometry of visual hull construction (as intersection of visual cones)

In the following section we review related previous work on visual hulls and probabilistic contour models. We then present our model for probabilistic image-based visual hulls. Next, we show an application of this approach to the reconstruction of pedestrian images. We demonstrate that the Bayesian reconstructions are more accurate than reconstructions based on the raw silhouettes. We conclude with a discussion and the avenues for future work.

2. Previous work

A visual hull (VH) is defined by a set of camera locations, the cameras' internal calibration parameters, and silhouettes from each view. Most generally, it is the maximal volume that creates all possible silhouettes of an object. The VH is known to include the object, and to be included in the object's convex hull. In practice, the VH is usually computed with respect to a finite, often small, number of silhouettes. (See Figure 1.) One efficient technique for generating the VH computes the intersection of the viewing ray from each designated viewpoint with each pixel in that viewpoint's image [14]. A variant of this algorithm approximates the surface of the VH with a polygonal mesh [13]. See [11, 13, 14] for the details of these methods.

While we restrict our attention to visual hulls from calibrated cameras, recent work has shown that visual hulls can be computed from weakly calibrated or uncalibrated views [18, 25]. Detailed models can be constructed from visual hulls with view-dependent reflectance or texture and accurate modeling of opacity [15].

A traditional application of visual hulls is the creation of models for populating virtual worlds, either for detailed models computed offline using many views (perhaps acquired using a single camera and turntable), or for online acquisition of fast and approximate models for real-time interaction. Visual hulls can also be used in recognition applications. Recognition can be performed directly on 3-D visibility structures from the visual hull (especially appropriate for the case of orthogonal virtual views), or the visual hull can be used in conjunction with traditional 2-D recognition algorithms. In [20] a system was demonstrated which rendered virtual views of a moving pedestrian for integrated face and gait recognition using existing 2-D recognition algorithms.

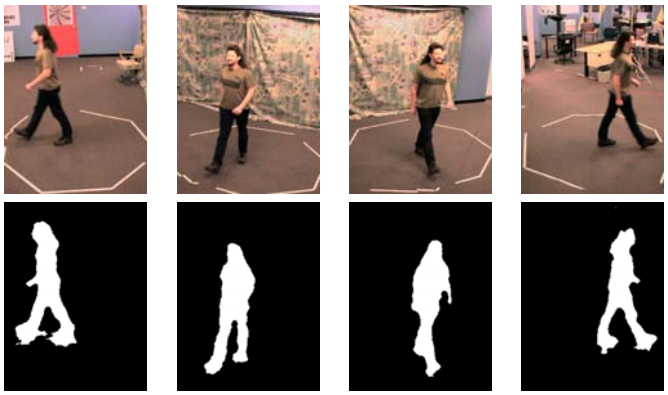
In this paper we consider visual hulls constructed from closed contours of pedestrian images. The authors of [1] developed a single-view model of pedestrian contours, and showed how a linear subspace model formed from principal components analysis (PCA) could represent and track a wide range of motion [2]. The Active Shape Model of [5] used a similar technique and was successfully applied to model facial variation.

The use of linear manifolds estimated by PCA to represent an object class, and more generally an appearance model, has been developed by several authors [4, 8, 23]. A probabilistic interpretation of PCA-based manifolds has been introduced in [6, 24] as well as in [16], where it was applied directly to face images. Snakes [9] and Condensation (particle filtering) [7] have also been used to exploit prior knowledge while tracking single contours.

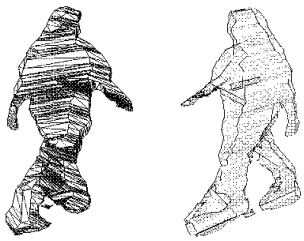
While regularization or Bayesian *maximum a posteriori* (MAP) estimation of single-view contours has received considerable attention as described above, relatively little attention has been given to multi-view data from several cameras simultaneously observing an object. With multi-view data, a probabilistic model and MAP estimate can be computed on implicit 3-D structures. In this paper we apply a PPCA-based probability model to form Bayesian estimates of multi-view contours used for visual hull reconstruction.

3. Bayesian image-based visual hulls

In this work, we derive a multi-view contour density model for 3-D visual hull reconstruction. We represent the silhouette shapes as sampled points on closed contours, with the shape vectors for each view concatenated to form a single



(a) Input



(b) Output

Figure 2: An example of VH data flow: (a) the input - a set of four images and the corresponding silhouettes; (b) the output - the reconstructed 3-D model, seen here from two different viewpoints.

vector in the input space. Our algorithm can be extended to a fixed number of distinguishable objects by concatenating their shape vectors, and to disconnected shapes more general than those representable by a closed contour if we adopt the level-set approach put forth in [12].

As discussed above, many authors have shown that a probabilistic contour model using PCA-based density models can be useful for tracking and recognition. An appealingly simple technique is to approximate a shape space with a linear manifold [5]. In practice, it is often difficult to represent complex articulated structures using a single linear manifold.

Following [4, 22], we construct a density model using a mixture of Gaussians PPCA model that locally models clusters of data in the input space with probabilistic linear manifolds. We model the uncertainty of a novel observation and obtain a MAP estimate for the low-dimensional coordinates of the input vector, effectively using the class-specific shape prior to restrict the range of probable reconstructions.

In the following section we see that if the 3-D object can be described by linear bases, then an image-based visual hull representation of the approximate 3-D shape of that object should also lie on a linear manifold, at least for the case of affine cameras.

3.1. Multi-view observation manifolds

If the vector of observed contour points of a 3-D object resides on a linear manifold, then the affine projections of that shape also form a linear manifold. Assume we are given a 3-D shape defined by the set of n points resulting from a linear combination of $3n$-D basis vectors. That is, the $3n$-D shape vector

$$\mathbf{p} = (\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_n)^T$$

can be expressed as

$$\mathbf{p} = \sum_{j=1}^{M} a_j \mathbf{b}^j = \mathbf{B}\mathbf{a}^T \tag{1}$$

where $\mathbf{a} = (a_1, ...a_M)$ are the basis coefficients for the M 3-D bases $\mathbf{b}^j = (\mathbf{b}_1^j, \mathbf{b}_2^j, ..., \mathbf{b}_n^j)^T$, $\mathbf{b}_i^j$ is the vector with the 3-D coordinate of point i in basis vector j, and $\mathbf{B}$ is the basis matrix whose columns are the individual $\mathbf{b}^j$ vectors. A matrix whose columns are a set of observed 3-D shapes will thus have rank less than or equal to M. Note that the coefficients $\mathbf{a}$ are computed for each given $\mathbf{p}$.

When a 3-D shape expressed as in (1) is viewed by a set of K affine cameras with projection matrices $\mathbf{M}_k$, we will obtain a set of image points which can be described as

$$\mathbf{c}_k = (\mathbf{x}_1^k, \mathbf{x}_2^k, ..., \mathbf{x}_n^k), \quad 1 \leq k \leq K, \tag{2}$$

where

$$\mathbf{x}_i^k = \mathbf{M}_k \mathbf{p}_i = \mathbf{M}_k \sum_{j=1}^{M} a_j \mathbf{b}_i^j = \sum_{j=1}^{M} a_j \mathbf{M}_k \mathbf{b}_i^j.$$

Therefore, $\mathbf{c}_k$ itself belongs to a linear manifold in the set of projected bases in each camera:

$$\mathbf{c}_k = \sum_{j=1}^{M} a_j \mathbf{q}_k^j = \mathbf{a}\mathbf{q}_k, \tag{3}$$

where $\mathbf{q}_k^j$ is the projected image of 3-D basis $\mathbf{b}^j$ in camera k:

$$\mathbf{q}_k^j = (\mathbf{M}_k \mathbf{b}_1^j, \mathbf{M}_k \mathbf{b}_2^j, ...\mathbf{M}_k \mathbf{b}_n^j)^T.$$

A matrix whose columns are a set of observed 2-D points will thus have rank less than or equal to M. For the construction of (1) - (3), we assume an ideal dense sampling of

points on the surface. The equations hold for the projection of all points on that surface, as well as for any subset of the points. If some points are occluded in the imaging process, or we only view a subset of the points (e.g., those on the occluding contour of the object in each camera view), the resulting subset of points can still be expressed as in (3) with the appropriate rows deleted. The rank constraint will still hold in this reduced matrix.

It is clear from the above discussion that if the observed points of the underlying 3-D shape lie on an M-dimensional linear manifold, then the concatenation of the observed points in each of the K views

$$\mathbf{o}_n = (\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_K)^T$$

can also be expressed as a linear combination of similarly concatenated projected basis views $\mathbf{q}_k^j$. Thus an observation matrix constructed from multiple instances of $\mathbf{o}_n$ will still be at most rank M.

3.2. Contour likelihood and prior

We should thus expect that when the variation in a set of 3-D objects is well-approximated by a linear manifold, their multi-view projection will also lie on a linear manifold of equal or lower dimension. When this is the case, we can approximate the density using PPCA with a single Gaussian. For more general object classes, object variation may only locally lie on a linear manifold; in these cases a mixture of manifolds can be used to represent the shape model [4, 22].

We construct a density model using a mixture of Gaussians PPCA model that locally models clusters of data in the input space with probabilistic linear manifolds. An observation is the concatenated vector of sampled contour points from multiple views. Each mixture component is a probability distribution over the observation space for the true underlying contours in the multi-view image. Parameters for the C components are determined from the set of observed data vectors $\mathbf{o}_n$, $1 \leq n \leq N$, using an EM algorithm to maximize a single likelihood function

$$L = \sum_{n=1}^{N} \log \sum_{i=1}^{C} \pi_i p(\mathbf{o}_n | i) \tag{4}$$

where $p(\mathbf{o}|i)$ is a single component of the mixture of Gaussians PPCA model, and π_i is the i^{th} component's mixing proportion. A separate mean vector μ_i, principal axes $\mathbf{W}_i$, and noise variance σ_i are associated with each of the C components. As this likelihood is maximized, both the appropriate partitioning of the data and the respective principal axes are determined. We used the Netlab [17] implementation of [22] to estimate the PPCA mixture model.

The mixture of probabilistic linear subspaces constitutes the prior density for the object shape. All of the images in the training set are projected into each of the subspaces associated with the mixture components, and the resulting means μ_i^t and covariances $\mathbf{\Sigma}_i^t$ of those projected coefficients are retained. The prior density is thus defined as a mixture of Gaussians, $P(\mathbf{P}) = \sum_{i=1}^{C} \pi_i N(\mu_i^t, \mathbf{\Sigma}_i^t)$.

The projection $\mathbf{y}$ of observation $\mathbf{o}_n$ is defined as a weighted sum of the projections into each mixture component's subspace,

$$\mathbf{y} = \sum_{i=1}^{C} p(i|\mathbf{o}_n)(\mathbf{W}_i{}^T(\mathbf{o}_n - \mu_i)), \tag{5}$$

where $p(i|\mathbf{o}_n)$ is the posterior probability of component i given the observation. To account for camera noise or jitter, we model the observation likelihood as a Gaussian distribution on the manifold with mean $\mu_o = \mathbf{y}$ and covariance $\mathbf{\Sigma}_o$: $P(\mathbf{o}|\mathbf{P}) = N(\mu_o, \mathbf{\Sigma}_o)$, where $\mathbf{P}$ is the shape.

Applying Bayes rule, we see that

$$P(\mathbf{P} = \mathbf{y} \mid \mathbf{o}) \;\propto\; P(\mathbf{o} \mid \mathbf{P} = \mathbf{y})\, P(\mathbf{P} = \mathbf{y}).$$

Thus the posterior density is the mixture of Gaussians that results from multiplying the Gaussian likelihood and the mixture of Gaussians prior:

$$P(\mathbf{P} = \mathbf{y} \mid \mathbf{o}) \propto \sum_{i=1}^{C} \pi_i N(\mu_i^p, \mathbf{\Sigma}_i^p). \tag{6}$$

By distributing the single Gaussian across the mixture components of the prior, we see that the components of the posterior have means and covariances

$$\begin{aligned} \mathbf{\Sigma}_i^p &= (\mathbf{\Sigma}_i^{t\,-1} + \mathbf{\Sigma}_o^{-1})^{-1}, \\ \mu_i^p &= \mathbf{\Sigma}_i^p \mathbf{\Sigma}_i^{t\,-1} \mu_i^t + \mathbf{\Sigma}_i^p \mathbf{\Sigma}_o^{-1} \mathbf{y}. \end{aligned} \tag{7}$$

The modes of this function are then found using a fixed-point iteration algorithm as described in [3]. The maximum of these modes, $\mathbf{x}^*$, corresponds to the MAP estimate, i.e., the most likely lower-dimensional coordinates in the subspace for our observation given the prior [2]. It is backprojected into the multi-view image domain to generate the reconstructed silhouettes $\mathbf{S}$. The backprojection is a weighted sum of the MAP estimate multiplied by the PCA bases from each mixture component of the prior:

$$\mathbf{S} = \sum_{i=1}^{C} p(i|\mathbf{x}^*)(\mathbf{W}_i(\mathbf{W}_i{}^T\mathbf{W}_i)^{-1}\mathbf{x}^* + \mu_i). \tag{8}$$

[2] Note that for a single Gaussian PPCA model with prior $N(\mu_t, \mathbf{\Sigma}_t)$, the MAP estimate is simply

$$x^* = \left(\mathbf{\Sigma}_t^{-1} + \mathbf{\Sigma}_o^{-1}\right)^{-1} \left(\mathbf{\Sigma}_t^{-1}\mu_t + \mathbf{\Sigma}_o^{-1}\mathbf{y}\right).$$

By characterizing which projections into the subspace are most likely, we restrict the range of reconstructions to be more like that present in the training set. Our regularization parameter is $\mathbf{\Sigma}_o$, the covariance of the density representing the observation's PCA coefficients. It controls the extent to which the training set's coefficients guide our estimate.

4. Reconstruction of pedestrian images

We have applied our method to the dataset of pedestrian sequences used in [20]. The images consist of four monocular views from cameras located at approximately the same height about 45 degrees apart. The working space of the system is defined as the intersection of their fields of view, and a simple color background model allows the extraction of a silhouette from each viewpoint. The use of a basic background subtraction method results in rough segmentation; body parts are frequently truncated in the silhouettes where the background is not highly textured, or else parts are inaccurately distended due to common segmentation problems from shadows or other effects. (See Figure 2 for example images from the experimental setup.)

The goal is to improve segmentation in the multi-view frames by reconstructing problematic test silhouettes based on MAP estimates of their projections into the mixture of lower dimensionsional subspaces (see Section 3.1 and 3.2). The subspaces are derived from a separate, cleaner subset of the silhouettes in the dataset. When segmentation improvements are made jointly across views, we can expect to see an improvement in the 3-D approximation constructed by the visual hull.

We represent each view's silhouette as sampled points along the closed contour extracted from the original binary images. All contour points are normalized to a common translation and scale invariant input coordinate system as follows. First, each image coordinate of the contour points (x, y) is transformed to the coordinates (x_r, y_r), in order to make points relative to an origin placed at that silhouette's centroid (x_c, y_c).

$$(x_r,\ y_r) \ = \ (x - x_c,\ y - y_c).$$

Next, points are normalized by d, the median distance between the centroid and all the points on the contour:

$$(x_n,\ y_n) \ = \ (x_r/d,\ y_r/d).$$

Finally, each view's vector of contour points is resampled to a common vector length using nearest neighbor interpolation. Empirically, resample sizes around 200 points were found to be sufficient to represent contours originating from (240 x 320) images and containing on average 850 points. The concatenation of the K views' vectors forms the final input.

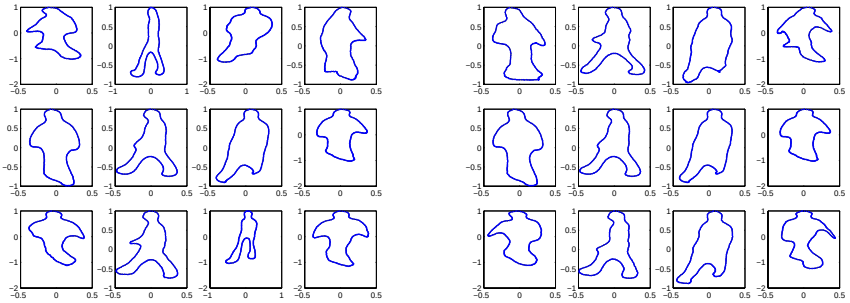
With the above alignments made to the data, inputs will still vary in two key ways: the absolute angle the pedestrian is walking across the system workspace, and the phase of their walk cycle at that frame. Unsurprisingly, we have found experimentally that reconstructions are poor when a single PPCA model is used and training is done with multi-view data from all possible walking directions and moments in gait cycle. Thus we group the inputs according to walking direction, and then associate a mixture of Gaussians PPCA model with each direction. Our visual hull system provides an estimate of the walking direction; however, without it we could still do image-based clustering. A novel input is then reconstructed using MAP estimation, as described in Section 3.2. In Figure 3 we show the first two multi-view principal components recovered for one of the mixture components' linear subspaces.

According to the visual hull definition, missing pixels in a silhouette from one view are interpreted as absolute evidence that all the 3-D points on the ray corresponding to that pixel are empty, irrespective of information in other views. Thus, segmentation errors may have a dramatic impact on the quality of the 3-D reconstruction. In order to examine how well the reconstruction scheme we devised would handle this issue and improve 3-D visual hull approximations, we tested sets of views with segmentation errors due to erroneous foreground/background estimates. We also synthetically imposed gross errors to test how well our method can handle dramatic undersegmentations. Visual hulls are constructed from the input views using the algorithm in [14].

The visual hull models resulting from the reconstructed views are qualitatively better than those resulting from the raw silhouettes (see Figures 4, 5). Parts of the body which are missing in one input view do appear in the complete 3-D approximation. Such examples illustrate the utility of modeling the uncertainty of an observed contour. In order to quantitatively evaluate how well our algorithm eliminates segmentation errors, we obtained ground truth segmentations for a set of the multi-view pedestrian silhouettes by manually segmenting the foreground body in each view. We randomly selected 15 frames from our test set to examine in this capacity. The mean squared error per contour point for the raw silhouettes in our ground truthed test set was found to be approximately 30 pixels, versus 11 pixels for the reconstructed silhouettes. This analysis is preliminary but promising.

5. Conclusions

We have developed a Bayesian approach to visual hull reconstruction using an image-based representation of object shape. We show how the use of a class-specific prior in visual hull reconstruction reduces the effect of segmentation errors in the silhouette extraction process. In our repre-



(a) The first principal component

(b) The second principal component

Figure 3: Primary modes of variation for the multi-view contours. The columns correspond to the four views. The middle row shows the mean contour for each view. The top and the bottom show the result of negative and positive variation along (a) the first and (b) the second principal component for one component of the mixture of PPCA model. The positive and negative variations are proportional to the largest and smallest PCA coefficients present in the training set, respectively.

sentation, 3-D information is implicit in the joint observation of multiple contours from known viewpoints. We use a mixture of probabilistic principal components analyzers to model the multi-view contour prior density. Our method was applied to a dataset of pedestrian sequences, and improvements in the approximate 3-D models under various noise conditions were shown. We plan to further test our method to see if our model improves accuracy in applications that use a visual hull for view synthesis in recognition tasks. We will also consider extensions to the proposed shape model that would allow the inference of 3-D structure.

References

[1] A. Baumberg and D. Hogg. Learning flexible models from image sequences. In *Proceedings of European Conference on Computer Vision*, 1994.

[2] A. Baumberg and D. Hogg. An adaptive eigenshape model. In *British Machine Vision Conference*, pages 87–96, Birmingham, September 1995.

[3] M. Carreira-Perpinan. Mode-finding for mixtures of Gaussian distributions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1318–1323, November 2000.

[4] T. Cootes and C. Taylor. A mixture model for representing shape variation. In *British Machine Vision Conference*, 1997.

[5] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, January 1995.

[6] J. Haslam, C. Taylor, and T. Cootes. A probabilistic fitness measure for deformable template models. In *British Machine Vision Conference*, pages 33–42, York, England, September 1994.

[7] M. Isard and A. Blake. Condensation – conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28, 1998.

[8] M. Jones and T. Poggio. Multidimensional morphable models. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 683–688, New Delhi, January 1998.

[9] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active shape models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[10] K. Kutulakos and S. Seitz. A theory of shape by space carving. In *Proceedings of the 7th IEEE International Conference on Computer Vision*, pages 307–314, Los Alamitos, CA, September 1999.

[11] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2):150–162, February 1994.

[12] M. Leventon, W. E. L. G. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 316–323, Hilton Head Island, SC, June 2000.

[13] W. Matusik, C. Buehler, and L. McMillan. Polyhedral visual hulls for real-time rendering. In *Proceedings of EGWR-2001*, pages 115–125, London, England, June 2001.

[14] W. Matusik, C. Buehler, R. Raskar, S. Gortler, and L. McMillan. Image-based visual hulls. In Kurt Akeley, editor, *Siggraph 2000, Computer Graphics Proceedings*, Annual Conference Series, pages 369–374. ACM Press / ACM SIGGRAPH / Addison Wesley Longman, 2000.

[15] W. Matusik, H. Pfister, A. Ngan, P. Beardsley, R. Ziegler, and L. McMillan. Image-based 3D photography using opacity hulls. In *Proceedings of the 29th Conference on Computer*

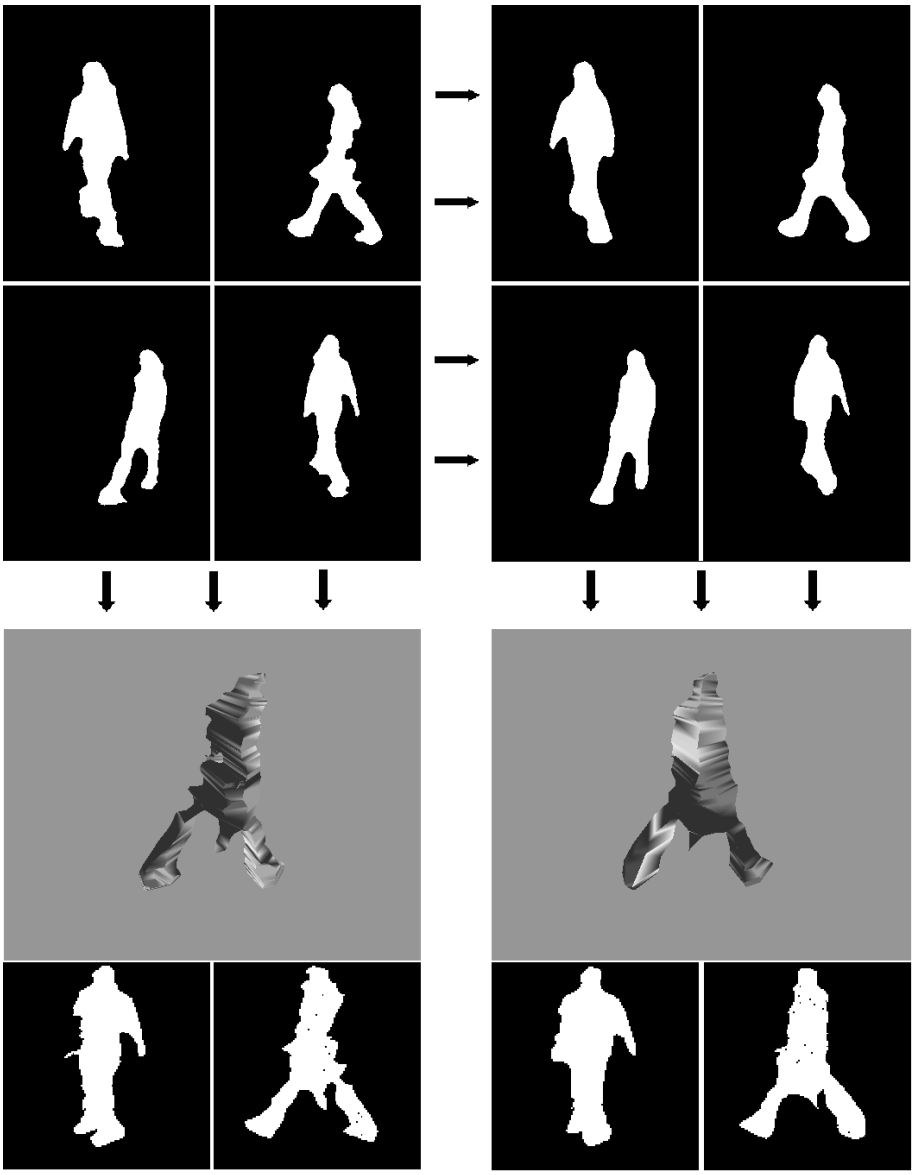


Figure 4: An example of segmentation improvement with PPCA-based Bayesian reconstruction. The four top-left images show the multi-view input, corrupted by segmentation noise. The four images directly to their right show the corresponding Bayesian reconstructions. The visual hull (VH) model is shown under the silhouettes from which it was constructed, for both the raw input (left) and the reconstructed silhouettes (right). Finally, virtual frontal and profile views projected from the VHs are shown at the bottom. Note how the right arm is missing in the virtual frontal view produced by the raw VH (bottom, leftmost image), whereas the arm is present in the Bayesian reconstructed version (bottom, image second from right).

Graphics and Interactive Techniques, ACM Transactions on Graphics, pages 427–437, New York, July 2002.

[16] B. Moghaddam. Principal manifolds and probabilistic subspaces for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):780–788, June 2002.

[17] Netlab. *http://www.ncrg.aston.ac.uk/netlab/index.html*.

[18] E. Boyer S. Lazebnik and J. Ponce. On computing exact visual hulls of solids bounded by smooth surfaces. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 156–161, Lihue, HI, December 2001.

[19] S. Seitz and C. Dyer. Photorealistic scene reconstruction by voxel coloring. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1067– 1073, San Juan,

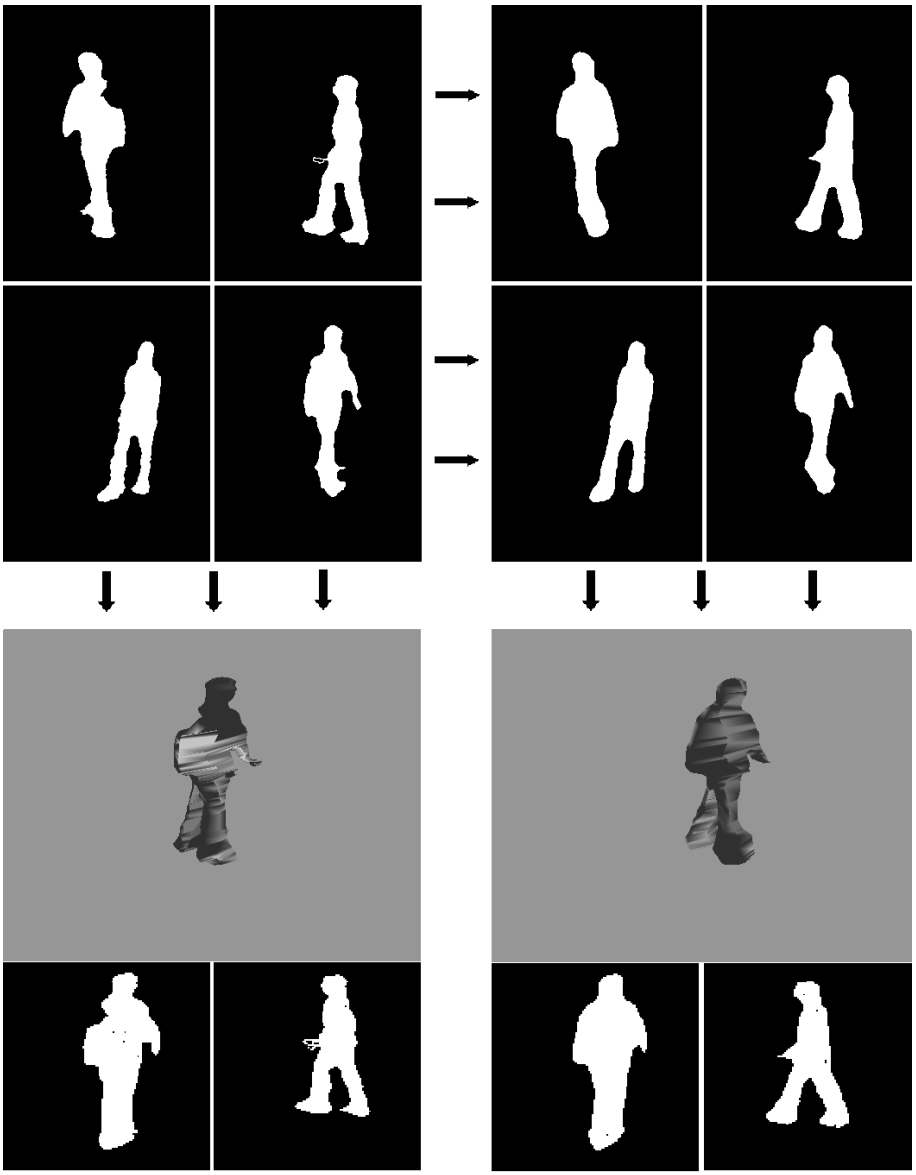


Figure 5: Another example of segmentation improvement with PPCA-based reconstruction. Please refer to the caption of Figure 4 for explanation.

Puerto Rico, June 1997.

[20] G. Shakhnarovich, L. Lee, and T. Darrell. Integrated Face and Gait Recognition From Multiple Views. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, Lihue, HI, December 2001.

[21] D. Snow, P. Viola, and R. Zabih. Exact voxel occupancy with graph cuts. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 345–353, Hilton Head Island, SC, June 2000.

[22] M. Tipping and C. Bishop. Mixtures of probabilistic principal component analyzers. *Neural Computation*, 11(2):443–482, 1999.

[23] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 586–590, Hawai, June 1992.

[24] Y. Wang and L. H. Staib. Boundary finding with prior shape and smoothness models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7):738–743, 2000.

[25] K. Wong and R. Cipolla. Structure and motion from silhouettes. In *Proceedings of the International Conference on Computer Vision*, pages 217–222, Los Alamitos, CA, July 2001.