In Proceedings of the 2nd Workshop on Statistical Methods in Video Processing, in conjunction with ECCV, Prague, Czech Republic, May 2004.

Virtual Visual Hulls: Example-Based 3D Shape Inference from Silhouettes

Kristen Grauman, Gregory Shakhnarovich, and Trevor Darrell

Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
{kgrauman,gregory,trevor}@csail.mit.edu

Abstract. We present a method for estimating the 3D visual hull of an object from a known class given a single silhouette or sequence of silhouettes observed from an unknown viewpoint. A non-parametric density model of object shape is learned for the given object class by collecting multi-view silhouette examples from calibrated, though possibly varied, camera rigs. To infer a 3D shape from a single input silhouette, we search for 3D shapes which maximize the posterior given the observed contour. The input is matched to component single views of the multi-view training examples. A set of viewpoint-aligned virtual views are generated from the visual hulls corresponding to these examples. The most likely visual hull for the input is then found by interpolating between the contours of these aligned views. When the underlying shape is ambiguous given a single view silhouette, we produce multiple visual hull hypotheses; if a sequence of input images is available, a dynamic programming approach is applied to find the maximum likelihood path through the feasible hypotheses over time. We show results of our algorithm on real and synthetic images of people.

1 Introduction

Estimating the 3D shape of an object is an important vision problem, with numerous applications in areas such as virtual reality, image-based rendering, or view-invariant recognition. Visual hull methods, also called Shape-From-Silhouette (SFS), yield general and compact shape representations, approximating the 3D surface of an object by intersecting the viewing cones formed by the rays passing through the optical centers of a set of cameras and their corresponding image silhouettes. Typically a relatively small number of input views (4-8) is sufficient to produce a compelling 3D model that may be used to create virtual models of objects and people in the real world, or to render new images for view-dependent recognition algorithms.

In the absence of calibrated cameras, Structure-From-Motion (SFM) techniques may be used with a sequence of data to estimate both the observed object's shape as well as the motion of the camera observing it. Most such algorithms rely on establishing point or line correspondences between images and frames, yet smooth surfaces without a prominent texture and wide-baseline cameras make correspondences difficult and unreliable to determine. Moreover, in the case of SFS, the occluding contours of the object are the only feature available to register the images. Current techniques for 3D reconstruction from silhouettes with an uncalibrated camera are constrained to the cases where the camera motion is of a known type, and the SFM methods cannot handle deformable, articulated objects.

In this paper we show that for shapes representing a particular object class, visual hulls (VHs) can be inferred from a single silhouette or sequence of silhouettes. Object class knowledge provides additional information about the object's structure and the covariate behavior of its multiple views. We develop a probabilistic method for estimating the VH of an object of a known class given only a single silhouette observed from an unknown viewpoint, with the object at an unknown orientation (and unknown articulated pose, in the case of non-rigid objects). We also develop a dynamic programming method for the case when sequential data is available, so that some ambiguities inherent in silhouettes may be eliminated by incorporating information revealed by how the object or camera moves.

We develop a non-parametric density model of the 3D shape of an object class based on many multi-view silhouette examples. The camera parameters corresponding to each multi-view training instance are known, but they are possibly different across instances. To infer a single novel silhouette's VH, we search for 3D shapes with maximal posterior probability given the observed contour. We use a nearest neighbor-based similarity search: examples which best match the contour in a single view are found in the database, and then the shape space around those examples is searched for the most likely underlying shape. Similarity between contours is measured with the Hausdorff distance. An efficient parallel implementation allows us to search 140,000 examples in a modest time.

To enable the search in a local neighborhood of examples, we introduce a new virtual view paradigm for interpolating between neighboring VH examples. Examples are re-rendered using a canonical set of virtual cameras; interpolation between 3D shapes is then a linear combination in this multi-view contour space. This technique allows combinations of VHs for which the source cameras vary in number and calibration parameters. The process is repeated to find multiple peaks in the posterior when the shape interpretation is ambiguous.

Our approach enables 3D surface approximation for a given object class with only a single silhouette view and requires no knowledge about either the object's orientation (or articulation), or the camera position. Our method can use sequential data to resolve ambiguities, or alternatively it can simply return a set of confidence-rated hypotheses (multiple peaks of the posterior) for a single frame. We base our non-parametric shape density model on the concise 3D descriptions that VHs provide: we can match the multi-view model in one viewpoint and then generate on demand the necessary virtual silhouette views from the training example's VH. Our method's ability to use multi-view examples from different camera rigs allows training data to be collected in a variety of real and synthetic environments.

2 Related Work

In this section we will review relevant related work on Shape-From-Silhouette algorithms and class-specific prior shape models.

Algorithms for computing the VH of an object have been developed based on the explicit geometric intersection of generalized cones [14]. Recent advances in VH construction techniques have included ways to reduce their computational complexity [16, 15], or to allow for weakly calibrated cameras [15]. A method combining SFS and stereo is given in [3] for the purpose of refining an object's VH by aligning its hulls from multiple frames over time. We rely on the efficient construction algorithm of [16] to calculate polygonal mesh VHs.

For the specific case of SFM using only the occluding contours, various methods have been devised which use knowledge of the uncalibrated camera's (or equivalently the rigid object's) motion to reconstruct surfaces. However, when applied to silhouette imagery current techniques are limited to certain types of camera motion (e.g., [4], [19], [21]).

When sequences of images are available, an alternative to geometric correspondence-based approaches like SFM is to utilize knowledge about the dynamics, or motion behavior, of the object moving in the video. For instance, knowledge about the types of motions the person is likely to perform may be exploited in order to infer the person's pose or shape. In the work of [2], a hidden Markov model is used to model the dynamics and 3D pose of a human figure in order to infer pose from a sequence of silhouettes by solving for the optimal path through the parametric model via entropy minimization. Our handling of sequential data uses dynamic programming to find the maximum likelihood path through a sequence of hypothesis virtual VHs. Our temporal integration step differs from that of [2] in that it processes different features (contours instead of central moments), and seeks to estimate a full 3D surface that fits the actual measurements of the input instead of rendering a cylinder-based skeleton from configural pose estimates.

A popular way to represent the variable shape of an object has been to employ a parametric distribution that captures the variation in the object shape. Such a model is often used for tracking, pose inference, or recognition. The use of linear manifolds estimated by PCA to represent an object class's shape, for instance, has been developed by several authors [13, 5, 1]. An implicitly 3D probabilistic shape model was introduced in [11], where a multi-view contour-based model using probabilistic PCA was given for the purpose of VH regularization. A method for estimating unknown 3D structure parameters with this model was given in [10]. However while [11, 10] require input views to be taken from cameras at the same relative angles as the training set, our method requires a single view with no calibration information at all.

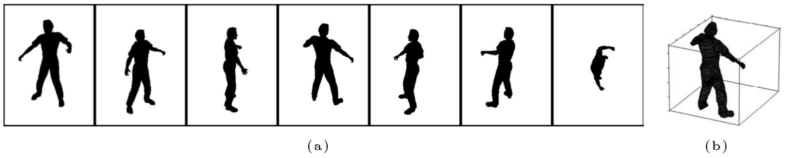


Fig. 1. (a) Examples in the model are composed of some number of silhouette views, plus their camera calibration parameters, which determine the VH (b).

Example-based and non-parametric density models of object shape have also been explored previously. In such models the object class is represented by a set of prototypical examples (or kernel functions centered on those examples), using either raw images or features extracted from them. For instance, the authors of [20] use 2D exemplars to track people and mouths in video sequences, a template hierarchy is developed for faster detection of pedestrian-shaped contours in [8], and in [17] a database of single view images with annotated body pose parameters is searched for a match to a test body shape based on its edges. In [18], a special mapping is learned from multi-view silhouettes to body pose. We employ a non-parametric density model constructed from many examples of multi-view silhouettes taken from different calibrated camera rigs.

We build a prior model of object shape using a non-parametric density model of multi-view contours. Given an observed single contour, we search for the shape most likely to have generated the observation. When the shape is ambiguous, we return multiple hypotheses corresponding to peaks in the posterior.

Our non-parametric density model for shape is defined by many multi-view silhouette instances of the object class, plus the associated camera calibration parameters (see Figure 1). Thus each example is equivalent to a VH. The number of cameras and their viewpoints may vary across examples, as long as for each example the camera parameters are recorded. The object's global orientation in each example is arbitrary. Generating a set of all possible examples to match discrete views is impractical, of course. A non-parametric density model gives us a principled way to interpolate between exemplars, which allows a relatively sparse set of examples to model a high dimensional space.

To measure the similarity between two contours A and B, represented by a set of uniformly sampled points, we use the Hausdorff distance:

$$||A - B||_H = \max(\max_{a \in A} D(a, B), \max_{b \in B} D(b, A)) \ , \qquad (1)$$

where $D(p, Q)$ is the shortest Euclidean distance from point p to any point in set Q. The Hausdorff distance has been proven to be an effective shape matching measure for such tasks as object recognition [12].

For the prior density over shapes we use a model of the form:

$$P(\mathbf{S}) = \frac{1}{Z} \sum_{i}^{N} K(\mathbf{S}, \mathbf{S}_i) \ , \tag{2}$$

where Z is a normalizing constant, $\mathbf{S}$ is the 3D shape, and K is a *kernel* function defined in terms of the distance between shapes. We define this distance D_r in terms of the two shapes' rendered appearance over all viewpoints:

$$D_r(\mathbf{S}, \mathbf{S}_i) = \int ||\mathbf{s}^{\mathbf{p}} - \mathbf{s}_i{}^{\mathbf{p}}||_H d\mathbf{p} \ , \tag{3}$$

where $\mathbf{s}^{\mathbf{p}}$ is the rendering of shape $\mathbf{S}$ from a camera at pose $\mathbf{p}$. For VHs constructed from a finite set of views, we approximate D_r using a set of m fixed camera locations:

$$D_r(\mathbf{S}, \mathbf{S}_i) \approx \sum_{\mathbf{j}=1}^{m} ||\mathbf{s}^{\mathbf{p}_j} - \mathbf{s}_i{}^{\mathbf{p}_j}||_H \ . \tag{4}$$

The training examples may be collected with real cameras, or generated synthetically (when the object class permits). The fact that the cameras in each example need not be the same is potentially a practical benefit when the model is built from real data, since this means that various camera rigs, at different locations, on different days, etc. may be employed to generate the examples.

2.1 3D Shape Inference with a Single Novel View

We will first explain the underlying method for estimating a single frame's VH. In fact, there could be several hypothesis estimates made at each frame in order to deal with ambiguous shapes; we discuss this process in Section 3.

To infer a VH from a single observed contour $\mathbf{C}$ in a Bayesian fashion, we wish to find $\mathbf{S}$ which maximizes

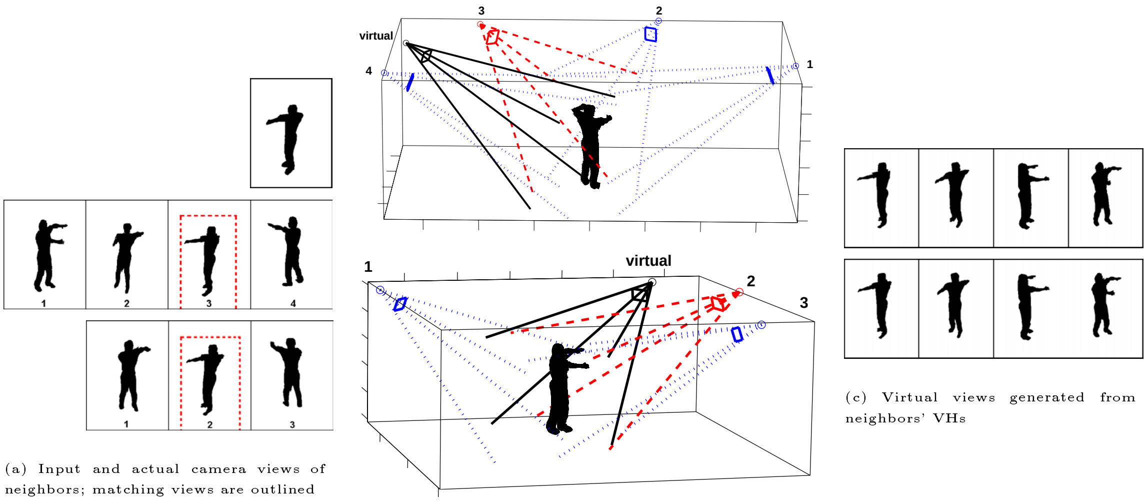
$$P(\mathbf{S}|\mathbf{C}) = \frac{P(\mathbf{C}|\mathbf{S})P(\mathbf{S})}{P(\mathbf{C})} \propto P(\mathbf{C}|\mathbf{S})P(\mathbf{S}) \ . \tag{5}$$

The observation likelihood is based on the similarity of a contour to any of the silhouette contours rendered from a given VH, which we approximate at a set of m discrete views at precomputed viewpoints:

$$P(\mathbf{C}|\mathbf{S}) = q \ \exp\left(-\min_{\mathbf{p}\in\Re^6} ||\mathbf{C} - \mathbf{s}^{\mathbf{p}}||_H^2 / 2\sigma^2\right) \approx \exp\left(-\min_{1<i<=m} ||\mathbf{C} - \mathbf{s}^{\mathbf{p}_i}||_H^2 / 2\sigma^2\right), \tag{6}$$

where q is a normalizing constant. Combining (2) and (6) we get the posterior that represents the most likely 3D shapes given an observed contour:

$$P(\mathbf{S}|\mathbf{C}) \propto \exp\left(-\min_{1<i<=m} ||\mathbf{C} - \mathbf{s}^{\mathbf{p}_i}||_H^2 / 2\sigma^2\right) \sum_{k=1}^{N} K\left(\sum_{j=1}^{m} ||\mathbf{s}^{\mathbf{p}_j} - \mathbf{s}_k^{\mathbf{p}_j}||_H\right). \tag{7}$$



(a) Input and actual camera views of neighbors; matching views are outlined

(b) Camera rigs of neighbors

(c) Virtual views generated from neighbors' VHs

Fig. 2. An example of rendering canonical views from the r-neighbors of the input. Views that matched the input are marked with dotted boxes (a), and their corresponding cameras in the two examples' rigs are also shown with dotted lines (b). Viewpoint-aligned virtual views generated from the two neighbors' VHs are in (c).

In practice, one may use any kernel K which vanishes for sufficiently large values of its argument. We approximate the sum in the last term in (7) by the sum over the examples $\mathbf{S}_{(1)}, \ldots, \mathbf{S}_{(N)}$ for which $\left\|\mathbf{C} - \mathbf{S}^{j}_{(k)}\right\|_{H} \leq r$, for some view j. These r*-neighbors* of the input contour $\mathbf{C}$ can be found by means of thresholded similarity search in the database with respect to the Hausdorff distance [1].

A key problem is how to combine multiple 3D shape examples to form a single shape estimate; naive interpolation of unaligned triangle meshes will not yield meaningful shapes. Instead, we propose to interpolate between VHs through weighted combinations of the multi-view contours rendered from a set of canonical viewpoints, defined as follows. Each r-neighbor $\mathbf{S}$ of the contour $\mathbf{C}$ has a particular stored view $\mathbf{S}^j$ that best matches $\mathbf{C}$. We first refine an example's matching viewpoint by searching closely around its best stored matching viewpoint to find the view $\mathbf{S}^{j\prime}$ that even better matches the input silhouette in terms of contour distance. The first canonical viewpoint $\mathbf{p}_1$ for $\mathbf{S}$ corresponds to the viewpoint of the camera $j\prime$. The second canonical view $\mathbf{p}_2$ is obtained by a fixed transformation applied to $\mathbf{p}_1$ (say, rotation around the centroid of $\mathbf{S}$'s VH by the angle θ), etc. for v canonical viewpoints. This allows us to render, for each $\mathbf{S}$, v virtual silhouettes $\mathbf{r}^1, \ldots, \mathbf{r}^v$ so that every $\mathbf{r}^j$ corresponds to similar viewpoints

[1] Similarity is measured on translation and scale invariant representations of the contours, which we obtain by subtracting the silhouette's 2D center of mass from each contour point's image coordinate and normalizing by its approximate size.

(relative to the shape configuration) for all the neighbors of $\mathbf{C}$. Even though the real world position and orientation of the camera from each example which saw a view similar to the input may differ, each virtual view will be viewpoint-aligned across the neighbor examples from which they were generated.

To clarify with an example: suppose two multi-view neighbors $\mathbf{S}_{(1)}$ and $\mathbf{S}_{(2)}$ for a novel input contain four and three views with camera parameters $\{\mathbf{p}_1^1, \ldots, \mathbf{p}_4^1\}$ and $\{\mathbf{p}_1^2, \ldots, \mathbf{p}_3^2\}$, respectively (see Figure 2). Suppose the third view in the first example and the second view in the second example matched the input. Then the first virtual view for the first example is taken from the projection of its VH onto the image plane corresponding to the virtual camera found by rotating $\mathbf{p}_3^1$ by θ degrees about the VH's vertical axis. Similarly, the first virtual view for the second example is taken from a camera placed θ degrees from $\mathbf{p}_2^2$. Subsequent virtual views for each example are taken at equal intervals relative to these initial virtual cameras. After taking a weighted combination of the contours from these aligned-viewpoint virtual views, the output VH will be constructed using the camera parameters of the nearest neighbor's similar view and its virtual cameras; since it was the best match for the input in a single view, it is believed to contain a viewpoint relative to the object that is most like the true unknown input camera's, up to a scale factor.

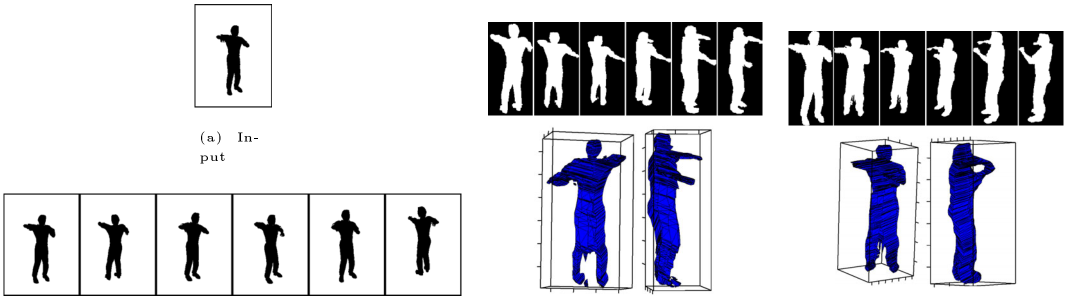
After the set of canonical contours is produced as described above, they are normalized in location (translated) and length (resampled) in order to align the contour points in the same view across all the neighbors of $\mathbf{C}$ so that they may be combined, per view. We would like to obtain the resulting shape as a linear combination of the neighbors; recall, however, the previously stated objective of maximizing the *a posteriori* probability of the shape. Thus, we must find the vector of weights $\mathbf{w}$ such that

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmax}} P\left(\sum_{k=1}^{N} w_k \mathbf{S}_{(k)} | \mathbf{C}\right) = \underset{\mathbf{w}}{\operatorname{argmax}} P\left(\mathbf{C} | \sum_{k=1}^{N} w_k \mathbf{S}_{(k)}\right) P\left(\sum_{k=1}^{N} w_k \mathbf{S}_{(k)}\right). \quad (8)$$

This can be done by means of gradient descent on the components of the vector $\mathbf{w}$. The shape hypothesis corresponding to the weights is then $\sum_k w_k^* \mathbf{S}_{(k)}$; note that this shape is, generally, no longer in the database, and may provide a better match for the input than any single training shape.

The 3D shape of a single view silhouette is inherently ambiguous: self-occlusions make it impossible to determine the full shape from a single frame, and the global orientation of the object may be uncertain if there is symmetry in the shape (e.g., a silhouette frontal view of a person standing with their legs side by side is similar to the view from behind). Thus we can expect the VHs corresponding to the "neighbor" single view silhouettes to manifest these different possible 3D interpretations. To combine widely varying contours from the neighbors' very different 3D shapes would produce a meaningless result (see Figure 3).

Instead, we maintain hypotheses corresponding to multiple peaks in the posterior at each time step. The nearest neighbors' aligned multi-view virtual silhouettes are clustered into enough groups such that the distance between two multi-view examples in one cluster is less than a threshold. Each cluster of examples will yield one peak of the posterior - one hypothesis VH. The single frame



(a) Input

(b) Similar component views from nearest neighbors

(c) Hypothesis 1

(d) Hypothesis 2

Fig. 3. An example where the input shape (a) has multiple interpretations. Its six nearest neighbors (b) contain views from examples originating from two general types of shapes: one front-facing body with the right elbow extended, the other back-facing with the left elbow extended. Aligned-viewpoint virtual views are projected from each of the six neighbors' VHs, and our algorithm finds two separate shape hypotheses (c,d) based on how these multi-view images cluster. Each hypothesis is shown from a frontal view and right side view, with the mean virtual views for each hypothesis shown above.

confidence in the VH hypothesis originating from a given cluster of neighbors is obtained by evaluating the posterior at the inferred VH for that cluster (see Eqn 7). Contours within the same cluster are topologically similar enough that they are expected to combine to form a valid shape. The hypotheses may be returned to a higher-level vision module, together with their confidences, or else they may be integrated using temporal data, as described below.

3 Integrating Single View Observations Over Time

When a sequence of observations is available, we apply a dynamic programming approach to find the maximum likelihood path through the feasible VHs at each time step, given our observations over time and the probabilities of transitioning from one shape to another. We construct a directed graph where each node is a time-indexed VH hypothesis (see Figure 4). Each directed path through the graph represents a legal sequence of states. Probabilities are assigned to each node and arc; node probabilities P_n^t are conditional likelihoods and arc probabilities P_a^t are transition probabilities:

$$\begin{aligned} P_n^t &= P(\mathbf{C}^t|\mathbf{S}^t = \mathbf{S}_i) \\ P_a^t &= P(\mathbf{S}^t = \mathbf{S}_j|\mathbf{S}^{t-1} = \mathbf{S}_i) \propto 1/D_r(\mathbf{S}_j, \mathbf{S}_i) \; . \end{aligned} \tag{9}$$

P_n^t is thus an estimate of the probability of observing contour $\mathbf{C}$ at time t given that the 3D shape of the object at time t is best approximated by the i^{th} cluster hypothesis's VH, $\mathbf{S}_i$, and this is the evaluation of the likelihood (see Eqn 6). P_a^t is a measure of the similarity between two hypotheses at consecutive

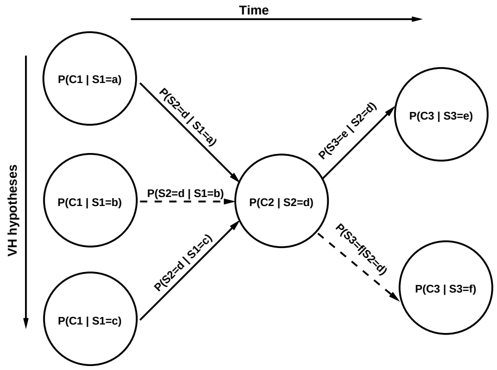


Fig. 4. Illustration of a directed graph corresponding to three consecutive frames. Nodes are VH hypotheses, edges are shape transition probabilities. The dotted line indicates the ML path found with dynamic programming.

time steps, which we estimate in our experiments in terms of the sum of the Hausdorff distances between a set of canonical views rendered from the respective hypothesis VHs.

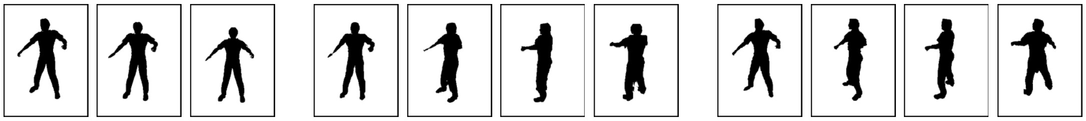
The maximum likelihood sequence of hypotheses in the graph is found using dynamic programming [7]. In this way an optimal path through the hypotheses is chosen, in that it constrains shapes to vary smoothly and favors hypotheses that were most likely given the observation. Note that this method for integrating temporal data remains general enough to handle different classes of objects, and requires no explicit model of dynamics for the object class. This process may be performed over windows of the data across time for longer sequences.

4 Experiments

We chose to build the model for human bodies from synthetic silhouettes using Poser, a computer graphics package [6]. The 3D person model is rendered from various viewpoints in randomized poses, and its silhouettes and camera parameters are recorded. An efficient parallel implementation allowed us to search 140,000 examples in modest time [2]. In our experiments we found the Hausdorff distance to be a robust measure of similarity between contours, since our input silhouettes were well-segmented, and we were able to reliably compute a scale and translation invariant representation of the contours.

We tested the inference method on both real and synthetic images. For the synthetic image tests, we generated a separate set of multi-view silhouettes using Poser, and then withheld each test example's VH information for ground truth

[2] More recently we have designed a fast contour matching technique based on approximate nearest neighbors and the Earth Mover's Distance that allows us to use more elaborate shape descriptors and query a database of the same size in 1.5 seconds on a single processor [9].



(a) Input view (left) and two nearest neighbors

(b) Views from inferred VH

(c) Same views from actual VH

Fig. 5. Example of ground truth comparison for test on synthetic input with $E = 73$.

comparisons. One view from each synthetic test example was input to our algorithm, and we measured the error E of the set of output VH hypotheses H for one test example as $E = \min_{h \in H} \left(\sum_{i=1}^{4} ||\hat{\mathbf{s}}^{\mathbf{p}_i} - \mathbf{s}^{\mathbf{p}_i}||_H \right)$, where $\hat{\mathbf{s}}^{\mathbf{p}_i}$ is the virtual view seen by a camera at pose $\mathbf{p}_i$ as inferred by our algorithm, and $\mathbf{s}^{\mathbf{p}_i}$ is the actual view seen by a camera at that pose for the withheld ground truth VH. For a synthetic set of 20 examples, the mean sum of Hausdorff distances over four views was 80. The Hausdorff distance errors are in the units of the image coordinates, thus a mean error of 80 summed over four views means that on average, per view, the farthest a contour point was from the ground truth contour (and vice versa) was 20 pixels. The synthetic test images are 240 x 320, with the contours having on average 800 points, and the silhouettes covering about 8,000 pixels in area. A typical example comparison between the virtual views generated by our estimated VH and the ground truth VH is shown in Figure 5.

We also inferred VHs for real images (see Figures 6 and 7). We consider these results to be preliminary but promising; the inferred VH hypotheses appear to provide a reasonable approximation of the 3D shape. The arm positions in Figure 7 are not optimal, and we plan to explore additional constraints on interpolation to improve the result.

5 Conclusions and Future Work

We developed a non-parametric prior model for a VH shape representation, and showed how 3D shape could be inferred from a single input silhouette or sequence of silhouettes with unknown camera viewpoints. Our prior model is learned by collecting multi-view silhouette examples from calibrated, though possibly varied, camera rigs. Visual hull inference consists of finding the shape hypotheses most likely to have generated the observed 2D contour. These peaks in the posterior are either returned directly, or, when a sequence of observations is available, integrated using a dynamic programming technique to find the most consistent trajectory of shapes over time.

Interpolation between neighboring examples allows our method to return shape estimates that are not literally in the set of examples used to define the prior model. We developed a new technique for 3D shape interpolation, using a set of viewpoint-aligned virtual views which are generated from the VHs corresponding to nearby examples. Interpolation between the contours of the aligned

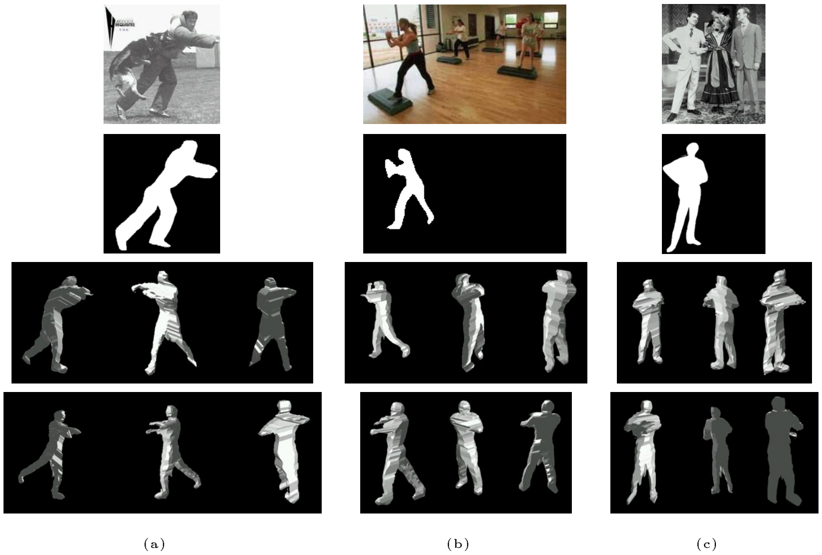


Fig. 6. Three example results on real single frame inputs. There are multiple hypothesis VHs due to the ambiguity in the single frame shapes (first two hypotheses are shown here for each example). Three viewpoints are rendered for each hypothesis.

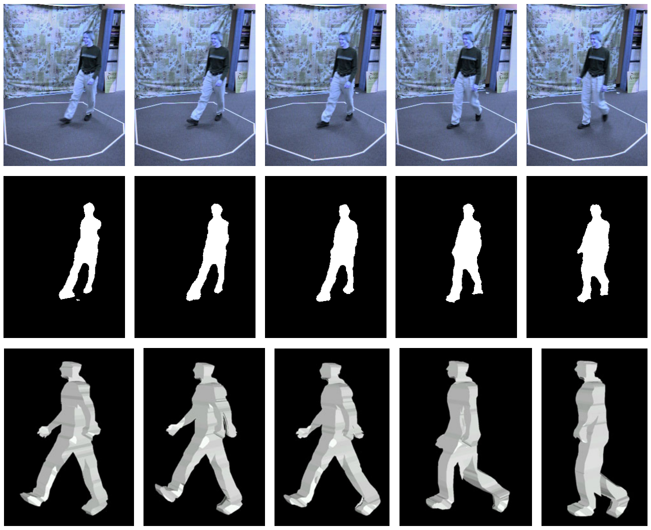


Fig. 7. Example result on real sequential data. Top row shows input sequence, middle row shows extracted silhouettes, and bottom row (output) shows VH hypotheses lying on the ML path, rendered here from a side view of the person in order to display the 3D quality of the estimate.

views produces a new set of silhouettes that are used to form the output VH approximating the 3D shape of the novel input.

We demonstrated our algorithm using a prior model trained with a large number of synthetic images of people. The accuracy of shape inference was evaluated quantitatively with held-out synthetic test images, and qualitatively with real images. We expect our method to be useful anytime a fast approximate 3D model must be acquired for a known object class, yet calibrated cameras or multiple cameras are not available. In the future we intend to explore ways to optimally combine the neighbors' contours, to deal with clutter in the shape matching stage, and to empirically study how the compactness of the database relates to the shape representational power of our method.

References

1. Baumberg, A., Hogg, D. An adaptive eigenshape model. BMVC, 1995.
2. Brand, M. Shadow puppetry. ICCV, 1999.
3. Cheung, G. K. M., Baker, S., Kanade, T. Shape-From-Silhouette of articulated objects and its use for human body kinematics estimation and motion capture. CVPR, 2003.
4. Cipolla, R., Blake, A. Surface shape from the deformation of apparent contours. IJCV 9(2) 1992.
5. Cootes, T., Taylor, C.A mixture model for representing shape variation.BMVC,1997.
6. Curious Labs, Egisys Co. Poser 5: The ultimate 3D character solution.
7. Forsyth, D., Ponce, J. Computer Vision: A Modern Approach. 2003. pp. 552–554.
8. Gavrila, D., Philomin, V. Real-time object detection for smart vehicles. ICCV 1999.
9. Grauman, K., Darrell, T. Fast contour matching using approximate earth mover's distance. CVPR, 2004.
10. Grauman, K., Shakhnarovich, G., Darrell, T. Inferring 3D structure with a statistical image-based shape model. ICCV, 2003.
11. Grauman, K., Shakhnarovich, G., Darrell, T. A Bayesian approach to image-based visual hull reconstruction. CVPR 2003.
12. Huttenlocher, D. Klanderman, G., and Rucklidge, W. Comparing images using the Hausdorff distance. PAMI, 1993.
13. Jones, M., Poggio, T. Multidimensional morphable models, ICCV, 1998.
14. Laurentini, A. The visual hull concept for silhouette-based image understanding. PAMI 16(2), 1994.
15. Lazebnik, S., Boyer, E., and Ponce, J. On computing exact visual hulls of solids bounded by smooth surfaces. CVPR, 2001.
16. Matusik, W., Buehler, C., McMillan, L. Polyhedral visual hulls for real-time rendering. EGWR, 2001.
17. Mori, G., Malik, J. Estimating human body configuration using shape context matching. ECCV, 2002.
18. Rosales, R., Sclaroff, S. Specialized mappings and the estimation of body pose from a single image. HUMO, 2000.
19. Szeliski, R., Weiss, R. Robust shape recovery from occluding contours using a linear smoother. IJCV 28(1), 1998.
20. Toyama, K., Blake, A. Probabilistic Exemplar-based tracking in a metric space. ICCV 2001.
21. Wong, K-Y. K. and Cipolla, R. Structure and motion from silhouettes. ICCV, 2001.