Using Strong Shape Priors for Stereo

Yunda Sun, Pushmeet Kohli, Matthieu Bray, and Philip H.S. Torr

Department of Computing,
Oxford Brookes University, UK
samy.sun@mail.edu.cn,
{pushmeet.kohli, philiptorr}@brookes.ac.uk
http://cms.brookes.ac.uk/research/visiongroup/

Abstract. This paper addresses the problem of obtaining an accurate 3D reconstruction from multiple views. Taking inspiration from the recent successes of using strong prior knowledge for image segmentation, we propose a framework for 3D reconstruction which uses such priors to overcome the ambiguity inherent in this problem. Our framework is based on an object-specific Markov Random Field (MRF)[10]. It uses a volumetric scene representation and integrates conventional reconstruction measures such as photo-consistency, surface smoothness and visual hull membership with a strong object-specific prior. Simple parametric models of objects will be used as strong priors in our framework. We will show how parameters of these models can be efficiently estimated by performing inference on the MRF using dynamic graph cuts [7]. This procedure not only gives an accurate object reconstruction, but also provides us with information regarding the pose or state of the object being reconstructed. We will show the results of our method in reconstructing deformable and articulated objects.

1 Introduction

Obtaining 3D reconstructions of objects from multiple images is a fundamental problem in computer vision. Reflecting the importance of the problem, a number of methods have been proposed for its solution. These range from methods such as shape from silhouettes [14] and space carving [11] to image based methods [12]. However, the problem of obtaining accurate reconstructions from sparse multiple views still remains far from being solved. The primary problem afflicting reconstruction methods is the inherent ambiguity in the problem (as shown in figure 1(a)) which arises from the many-one nature of the mapping that relates 3D objects and their images.

Intuitively the ambiguity in the object reconstruction can be overcome by using prior knowledge. Researchers have long understood this fact and weak priors such as surface smoothness have been used in a number of methods [8,13,15]. Such priors help in recovering from the errors caused by noisy data. Although they improve results, they are weak and do not carry enough information to guarantee a unique solution. At this point, the question to be asked is: *Can we make use of stronger prior knowledge?* A possible source for a strong prior could be *the knowledge of the shape of the object we are trying to reconstruct*. In other words, if we know which object we are trying to reconstruct, we can use a strong *object-specific* prior to force the reconstruction to look like that object.

P. Kalra and S. Peleg (Eds.): ICVGIP 2006, LNCS 4338, pp. 882–893, 2006.

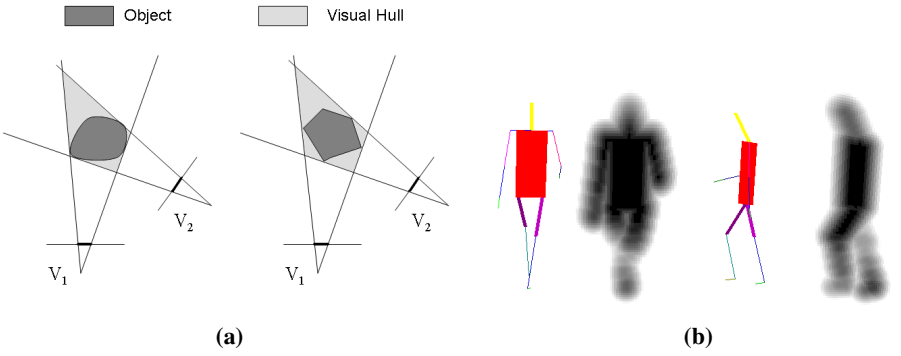

Fig. 1. a) Ambiguity in object reconstruction from sparse multiple views. The figure shows how two completely different objects can have the same visual hull. Further, if both objects have the same colour, the photo hull and their projections on multiple viewpoints would also be the same. b) Example of an articulated model. The figure shows a simple stick-model of a human in different poses and the corresponding priors as its 3D distance transforms used to set up our energy described in Section 2.

Strong Object-Specific Priors. Kumar *et al.* [10] proposed a method for using strong priors for solving image segmentation. They introduced the "Object-Specific Markov Random Field" model which combined Markov Random Fields (MRFs) with an *object-specific* shape prior. This shape-prior was defined by a Layered Pictorial Structures (LPS) model. The LPS model provided them with a strong prior able to model shape variations parameterized by a set of latent shape parameters. They obtained good object localization and segmentation results using their approach. However, their method required a large library of exemplars for different parts for the LPS model. Bray *et al.* [4] suggested using a simple articulated model. This makes the problem easier to solve computationally while still giving excellent segmentation results.

Parametric Models of Strong Prior Knowledge. In this work we will investigate the use of parametric models of objects as strong priors on the reconstruction, together with the weak prior of surface smoothness. The models are parameterized by a set of latent shape parameters which inherently characterize the state of the object to be reconstructed. Specifically we illustrate our ideas in terms of two model categories: articulated and deformable. Models belonging to the first category are used as strong priors for reconstructing articulated objects such as humans. They are parameterized by a set of pose parameters which characterize the pose of the object. Figure 1(b) shows an example of an articulated human model. Models belonging to the second category are used as strong priors on active-shape or deformable objects. The individual instances of these objects might be different from each other but they can be described by a common high-level parametrization. For example, objects like chairs can be parameterized in terms of parameters like height, width of seat etc. A deformable model for a vase is shown in figure 3(a).

Framework for Integrating Strong Prior Knowledge. A Bayesian approach to solve the 3D stereo reconstruction problem would typically be to formulate it in terms of a MRF.

This offers us the advantage of a seamless integration of strong priors (as defined above) with data, in this case conventional reconstruction measures such as photo-consistency, surface smoothness and visual hull membership. Inference on the random variables constituting the MRF can be seen as an energy minimization problem. If this energy function is *regular* (explained in Section 3) then its solution can be obtained in polynomial time using efficient graph-cut algorithms [9].

Inference of Model Shape Parameters. To guarantee an object-like reconstruction, our prior should have latent variables that model the shape variability of our object of interest. Then we optimize the energy of the *object-specific* MRF with respect to all these latent variables. Thus obtaining at the same time an accurate reconstruction as well as an estimate of the latent parameters. As explained in section 3, such an optimization procedure is extremely computationally expensive since it requires a graph cut to be computed multiple number of times. While performing this inference procedure, we make the observation that as we optimize over the model parameters, the energy function of the MRF we were trying to minimize changes minimally. This motivates us to use the recently proposed dynamic graph cut algorithm [7], which enables fast minimization of regular energy functions which change minimally from one instance to the next.

Organization of the Paper. The outline of the paper is as follows. We start by describing the object-specific MRF which forms the basis of this work. We explain how recently proposed methods for reconstruction can be explained in terms of this framework. The details of the efficient algorithm for performing inference over this MRF is given in section 3. In section 4, we will illustrate the use of this framework in reconstructing deformable and articulated objects, and provide results of experiments performed on real data. The conclusions and directions for future research are given in section 5.

2 Bayesian Framework

Within this section we provide a Bayesian formulation of the object reconstruction problem. This framework allows for the integration of strong object-specific priors with widely used data based terms such as photo-consistency and visual hull membership. We will also show how existing methods for object reconstruction such as [8,13,15] can be explained in this framework.

Object-Specific Markov Random Field for Reconstruction. A MRF comprises of a set of discrete random variables $\{X_1, X_2, \ldots, X_n\}$ defined on the index set $\mathcal{V}$, such that each variable X_v takes a value x_v from the label set $\mathcal{X} = \{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_l\}$ of all possible labels. We represent the set of all variables $x_v, \forall v \in \mathcal{V}$ by the vector $\mathbf{x}$. Unless noted otherwise, we use symbols i and j to denote values in $\mathcal{V}$. Further, we use $\mathcal{N}_v$ to denote the set consisting of indices of all variables which are neighbours of the random variable x_v in the graphical model.

For the reconstruction problem, the set $\mathcal{V}$ corresponds to the set of all voxels in the volume of interest, $\mathcal{N}$ is a neighbourhood defined on this set[1], the binary variable

[1] In this paper, we have used the standard 6-neighbourhood i.e. each voxel is connected to the 6 voxels surrounding it.

x_v denotes the labeling of the voxel $v \in \mathcal{V}$, and the set $\mathcal{X}$ comprises of two labels ('obj','empty') representing whether the voxel belongs to the empty space or not. We will use $\mathcal{H}$ to denote the set of all voxels present in the visual hull obtained from object silhouettes. Every configuration $\mathbf{x}$ of such an MRF defines a 3D object reconstruction.

Given a set of images $\mathbf{I}$ and (or) a visual hull $\mathbf{H}$ (obtained using silhouettes), collectively constituting the data $\mathbf{D}$, $\mathbf{D}$ could be images, measurements and it could also include the result of some other algorithm e.g. a visual hull, we wish to reconstruct a known object. This can be done by labelling each voxel v in the volume of interest $\mathcal{V}$ as belonging to the object reconstruction, or belonging to the scene. Taking a Bayesian perspective, the optimal labels for the voxels are those which maximize the posterior probability $p(\mathbf{x}|\mathbf{D})$, which can be written in terms of a Gibbs distribution as:

$$p(\mathbf{x}|\mathbf{D}) = \frac{p(\mathbf{D}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{D})} = \frac{1}{Z_{\mathbf{x}}} \exp(-\Psi(\mathbf{x})), \tag{1}$$

where $\Psi(\mathbf{x})$ is the energy of the configuration $\mathbf{x}$ of the MRF. The most probable or maximum a posteriori (MAP) reconstruction solution can be found by computing the least energy configuration $\mathbf{x}^* = \arg\min_{\mathbf{x}} \Psi(\mathbf{x})$. The energy $\Psi(\mathbf{x})$ corresponding to the configuration $\mathbf{x}$ consists of likelihood and prior terms. These can be written in terms of individual and pairwise interaction functions as:

$$\Psi(\mathbf{x}) = \sum_{i \in \mathcal{V}} (\psi(x_i) + \phi(\mathbf{D}|x_i) + \sum_j (\psi(x_i, x_j) + \phi(\mathbf{D}|x_i, x_j))) + \text{const.} \tag{2}$$

Specifying the Likelihood Terms. Given the data $\mathbf{D}$, the unary likelihood term $\phi(\mathbf{D}|x_i)$ specifies the penalty (or cost) for assigning the label x_i to the voxel v_i. Assuming $\mathbf{D} = \mathbf{H}$, we can define $\phi(\mathbf{D}|x_i)$ in terms of the visual hull as:

$$\phi(\mathbf{D}|x_i = \text{'obj'}) = \begin{cases} \alpha & \text{if } i \in \mathcal{H}, \\ \beta & \text{otherwise,} \end{cases} \tag{3}$$

where α and β are arbitrary constants and satisfy the property $\alpha < \beta$. Snow *et al.* [13] used raw images along with their binary segmentations to develop a generalized version of these terms. Their likelihood function incorporated the absolute difference in the intensities of the pixels which intersected at a voxel. Their approach can be viewed as using a visual hull where each voxel has an associated confidence value. In contrast to the above approach, Kolmogorov *et al.* [8] only used image information and assumed the segmentation to be unknown. They took ($\mathbf{D} = \mathbf{I}$) and used an image based photo-consistency measure to define $\phi(\mathbf{D}|x_i)$ as: $\phi(\mathbf{D}|x_i = \text{'obj'}) = \min\{0, (I_p - I_q)^2 - K\}$ where p and q are pixels in the images, which lie near the projection of the voxel i, and I_p and I_q are their intensities.

In their recent work on multi-view stereo, Vogiatzis *et al.* [15] took $\mathbf{D} = \{\mathbf{I} \text{ and } \mathbf{H}\}$ i.e. they used both the visual hull $\mathbf{H}$ and object images $\mathbf{I}$ as the data $\mathbf{D}$. They used a photo-consistency term that was obtained from the images. Further, instead of using the entire volume of interest, they only performed inference on the labels of voxels between two specific surfaces S_{base} and S_{in}. They defined S_{base} as the surface of the visual hull,

and defined S_{in} as the locus of voxels which are located at a specific distance d_{in} inside S_{base}. This is equivalent to using the unary likelihood term:

$$\phi(\mathbf{D}|x_i = \text{`obj'}) = \begin{cases} -\infty & \text{if} \quad i \in \mathcal{H}^-, \\ +\infty & \text{if} \quad i \notin \mathcal{H}, \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where $\mathcal{H}^-$ is the volume enclose by S_{in} and is in effect a contraction of the actual visual hull $\mathcal{H}$. Although the use of various measures for the unary likelihood have been investigated, the pairwise likelihood $\phi(\mathbf{D}|x_i, x_j)$ has remained relatively ignored by researchers. This term reflects the compatibility of two neighbouring latent variables in the MRF, and has been shown to be extremely useful in the context for the image segmentation problem, where it is called the contrast term [3,10]. We define this term as:

$$\phi(\mathbf{D}|x_i, x_j) = \lambda \exp\left(\frac{-g^2(i,j)}{2\sigma^2}\right) \frac{1}{dist(i,j)} \tag{5}$$

where $g^2(i,j)$ measures the difference in the estimated intensity values of the voxels i and j and $dist(i,j)$ gives the spatial distance between i and j. Such as estimate can be obtained either by using voxel colouring methods or directly from the object images in a manner analogous to the photo-consistency term. The effect of this term will be to favour discontinuities aligning with the object surface.

2.1 Incorporating Priors

We now describe how weak and strong prior information can be incorporated in our MRF framework.

Surface Smoothness as a Weak Prior. The pairwise interaction term $\psi(x_i, x_j)$ has been used in a number of methods as a weak prior to encourage smoothness in the reconstruction surface [8,13]. This is done by penalizing dissimilar label assignments in neighbouring voxels. The pairwise prior term takes the form of a Generalized Potts model:

$$\psi(x_i, x_j) = \begin{cases} K_{ij} \text{ if } x_i \neq x_j, \\ 0 \quad \text{if } x_i = x_j. \end{cases} \tag{6}$$

Parametric Models as Strong Priors. Suppose we know the object we are trying to reconstruct. Such information could be used to constrain the reconstruction result to look like the object and intuitively improve the reconstruction. However, we face two key problems at this juncture: (1) It is difficult to know what should be an appropriate representation for such knowledge. (2) How could we integrated such information in our Bayesian framework for the reconstruction problem? Our solution to the first problem is the use of generative parametric models to represent knowledge about the object. These models are parameterized by a set of parameters θ, which define the state of the object. The MRF formulation is shown in the graphical model shown in figure 2.

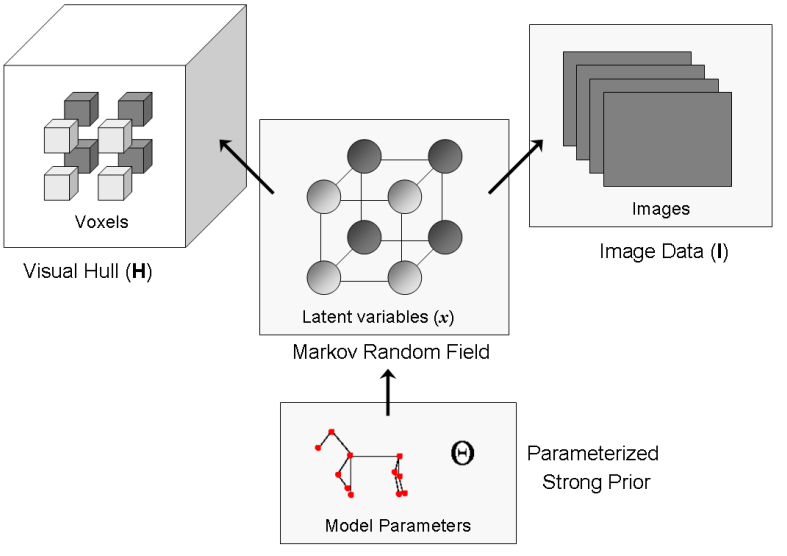

Fig. 2. The Bayesian framework for Object Reconstruction

In this framework the parameters of the object model are considered as latent (or hidden) variables. The energy function of the MRF is:

$$\Psi(\mathbf{x},\theta) = \sum_{i\in\mathcal{V}}(\psi(x_i|\theta) + \phi(\mathbf{D}|x_i) + \sum_{j}(\psi(x_i,x_j) + \phi(\mathbf{D}|x_i,x_j))). \tag{7}$$

For a particular value of θ, the model could be used to generate a coarse reconstruction of the object. This reconstruction is used to define the unary prior term $\psi(x_i|\theta)$. The function $\psi(x_i|\theta)$ is chosen such that given an estimate of the location and shape of the object, voxels near to that shape are more likely to be included in the reconstruction, the term used by us is: $\psi(x_i|\theta) = -\log p(x_i|\theta)$ where $p(x_i|\theta)$ is defined as:

$$p(x_i = \text{`obj'}|\theta) = \frac{1}{1+\exp(\mu * (d(i,\theta) - d_{sur}))} \tag{8}$$

where $d(i,\theta)$ is the distance of a voxel i from the surface generated by the parametric model and d_{sur} is the average distance from the model surface to the surface voxels in the true object reconstruction. The distance for all the voxels in the volume of interest is efficiently computed by performing a 3D distance transform [5]. An example of a 3D distance transform is shown in figure 1(b). The parameter μ determines the ratio of the magnitude of the penalty that points outside the shape prior have compared with points inside the shape.

3 MAP-MRF Inference Using Dynamic Graph Cuts

We next describe how to find the optimal configuration of the object specific MRF. As stated earlier this problem can be solved by minimizing the energy function defined by

the MRF. Energies like the one defined in (7) can be solved using graph cuts if they are *regular* [9]. In our case, this is indeed the case and thus for a particular value of θ, we can find the optimal configuration $\mathbf{x}^* = \min_{\mathbf{x}} \Psi(\mathbf{x}, \theta)$ using a single graph cut. The labels of the latent variable in this configuration give the optimal reconstruction.

3.1 Optimizing over the Parametric Model Parameters

Since our strong object-specific model prior is defined in terms of latent variables, we would like to make sure that it reflects the correct pose of the object. To do this we solve the problem: $\theta_{opt} = \arg\min_{\theta} \min_{\mathbf{x}} \Psi(\mathbf{x}, \theta)$. In our experiments we observed that the energy function projection $\Psi(\mathbf{x}^*, \theta)$ is locally uni-modal and can be optimized using standard techniques like gradient descent. The plots of this projection can be seen in figure 5(i). Our algorithm starts with an initial guess of the latent variables pose and optimizes it using standard minimization methods. Once an estimate of θ_{opt} has been found we can find the optimal reconstruction $\mathbf{x}_{opt} = \arg\min_{\mathbf{x}} \Psi(\mathbf{x}, \theta_{opt})$ using a single graph cut.

Minimizing Energies using Dynamic Graph Cuts. The minimization procedure for estimating θ_{opt} involves computing the value of $\min_{\mathbf{x}} \Psi(\mathbf{x}, \theta)$ for different values of θ. Each such computation requires a graph cut to be computed and if the time taken for computing this cut is high, it would make our optimization algorithm quite slow. Here we make the following observation: Between different iterations of the optimization algorithm, the change in the value of θ is small. This is reflected in the change in the energy function we are required to minimize, which is small as well. For such a sequence of energies, the graph cut computation can be made significantly faster by using the dynamic graph cut algorithm recently proposed in [7]. This algorithm works by using the solution of the previous graph cut computation for solving the new instance of the problem. In our experiments, we found that the dynamic algorithm was 15-25 times faster than the algorithm proposed in [2], which recomputes the st-mincut from scratch and has been shown to be the fastest algorithm for graphs commonly used in computer vision problems.

4 Applications

Within this section, we will show some results obtained by using the Bayesian framework defined in section 2. We apply our approach on two object categories to show how strong object-specific priors can help in obtaining accurate reconstructions from ambiguous and noisy data.

4.1 Deformable Models

Deformable models as the name suggests can alter their shape and in the process generate different instances of the object. These can be used while reconstructing objects with high intra class variability. The latent variables θ characterizing these models dictate the exact shape that the model takes. We illustrate their use in obtaining 3D reconstructions of a vase, from a few images shown in figure 3(b).

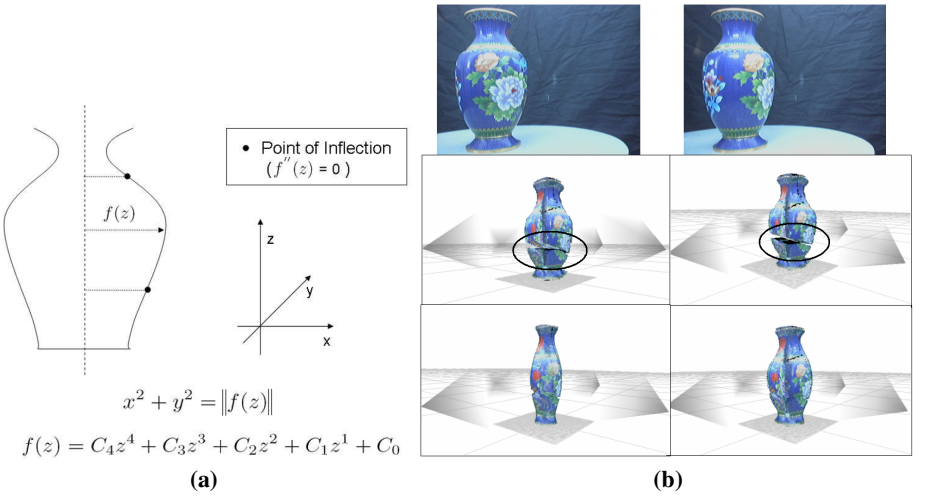

Fig. 3. a) The Parametric Deformable Vase Model. b) Object Reconstruction using Deformable Models. The images used for reconstruction are shown in row 1. The second row shows two views of the visual hull obtained using noisy silhouettes of the vase. In the third row, we show the results obtained by our method before/after optimizing the parameters of the deformable model. It should be noted here that the reconstruction results obtained by our method are smoother and do not suffer from discontinuities such as the cut seen in the visual hull.

The Parametric Vase Model. We use a rotationally symmetric model (shown in figure 3(a)) for the vase. The model is described in terms of circles in the horizontal plane as: $x^2 + y^2 = \|f(z)\|$ where $f(z)$ is a n-degree polynomial. In our experiments, we bound the degree of $f(z)$ to four, making it take the form:

$$f(z) = C_4z^4 + C_3z^3 + C_2z^2 + C_1z + C_0. \tag{9}$$

The coefficients $\{C_0, \ldots, C_4\}$ of the function constitute the set of latent parameters θ characterizing the shape of the model. We optimize over the values of these coefficients (as explained in section 3) to obtain a shape that acts as a coarse reconstruction of the actual object. The model can be strengthened by making it more object-specific. It can be observed that the vase surface has two inflection points. This constraint can be incorporated in our model by making sure that the second derivative of $f(z)$, which is defined as $f''(z) = 12C_4z^2 + 6C_3z + 2C_2$ has two unequal real roots. This gives us the constraint: $36{C_3}^2 - 96C_4C_2 > 0$.

Experiments. We use the images and silhouettes of the vase as data. These are obtained from four cameras which are uniformly distributed around the object as shown in figure 3(b). We quantize the volume of interest into 3×10^5 voxels. The *object-specific* MRF formulated for the reconstruction problem has 1.5×10^5 binary latent variables. The energy $\Psi(\mathbf{x}^*, \theta)$ of this MRF is constructed as described in section 2. We then minimize it with respect to the shape model parameters θ to obtain an estimate of θ_{opt}. The results from the experiment are shown in figure 3(b).

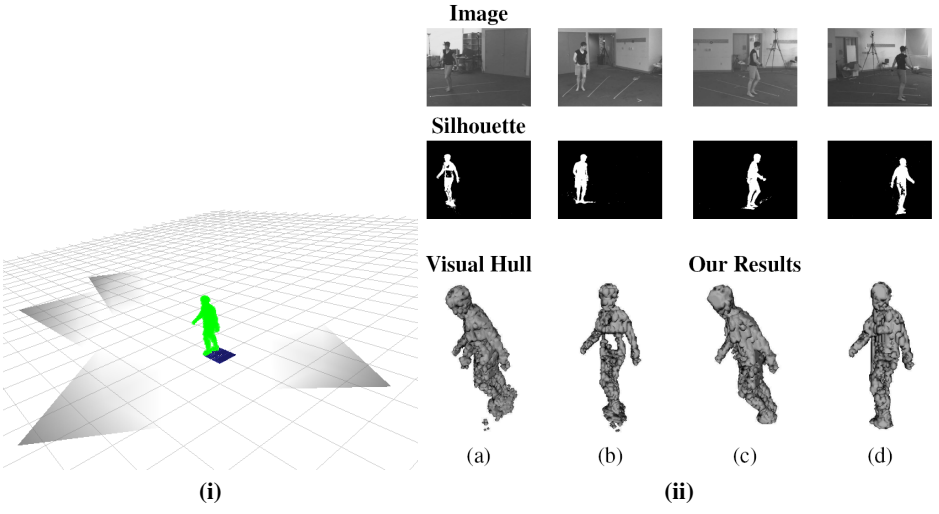

Fig. 4. i) Camera Positions and Reconstruction. The figure show the position and orientations of the four cameras which were used to obtain the images which constituted the data-set for our first experiment. We also see the reconstruction result generated by our method. ii) 3D Object Reconstruction using Strong Object-Specific priors. The first and second rows show the images and silhouettes used as the data. Two views of the visual hull generated using the data are shown in the first two columns of the bottom row ((a) and (b)). The visual hull is noisy and contain artifacts like the spurious third arm caused by the ambiguity in the problem. We are able to overcome such problems by using strong prior knowledge. The reconstructions obtained by our method are shown in column 3 and 4 ((c) and (d)).

4.2 Articulated Models

Articulated models not only help in reconstructing the object, but also provide information about its pose. In this section, we will use an articulated stick-man model to solve the challenging problem of reconstructing and estimating the shape (and pose) of humans. The problem is especially hard because humans have many joint angles and thus the parametric model needed to describe them will have a high number of latent variables.

The Stick-man model. We use a simple articulated stick-man model (shown in figure 1(b)) in our experiments to generate a rough pose-specific prior on the reconstruction of the human. The model is parameterized by a 26 dimensional pose vector θ that describes absolute position and orientation of the torso, and various other joint angle values. There are no constraints or joint-limits incorporated in our model.

Experiments. We use real and synthetic video sequences of humans as data. The data-set for our first experiment consists of videos sequences of four views of a human subject walking in a circle. This data-set is used in [1]. It comes with silhouettes of the human subject obtained using pixel-wise background intensity modeling. The cameras position and orientations with respect to the object are shown in the figure 4(i).

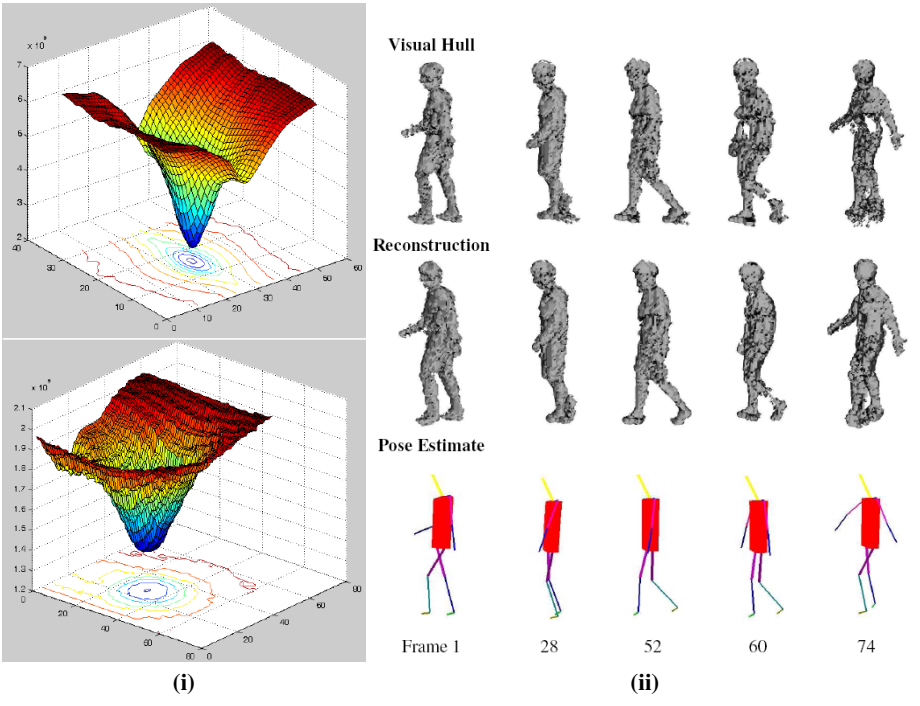

Fig. 5. i) The plots shows how the value of $\min_{\mathbf{x}} \Psi(\mathbf{x}, \theta)$ is affected by changes in the pose parameters of the stick model used to generate the reconstruction prior. The first plot shows the values obtained by varying the global translation and rotation parameters of the stick-man model in the x-axis. The second plot shows the values while varying the joint angles of the left shoulder in x and z axes. Observe that the effect of changing the joint angles of the left shoulder is less than the effect caused by changes in the global translation and rotation parameters. ii) Pose Inference and 3D Object Reconstruction results. The data-set is the same as used in [1] and consists of 4 views of a human subject walking in a circular path. Middle row: Reconstruction result. Bottom row: Pose estimate. Observe that we are able to get excellent reconstruction and pose estimation results even when the visual hull contains large errors (as seen in frame 60 and 74).

The first step in our method is the computation of the visual hull. The procedure starts with the quantization of the volume of interest as a grid of cubical voxels of equal size. Once this is done, each voxel center is projected into the input images. If any of the projections falls outside the silhouette, then the voxel is discarded. All remaining voxels constitute the visual hull. Some visual hull results are illustrated in figure 4(ii). It can be observed that because of the skewed distribution of cameras, the visual hull is quite different from the true object reconstruction. Further, as object segmentations are not accurate, it has large errors. The prominent defects in the visual hull results include: (i) The presence of holes because of segmentation errors in the object silhouettes (bottom row (b)), (ii) the presence of auxiliary parts caused by shadows, (iii) the *third-arm effect* resulting from self-occlusion and ambiguity in the reconstruction due to small number of views (bottom row (a)). It can be seen that our reconstruction results do not suffer from these errors (bottom row (c) and (d)).

Analysis of the Inference Algorithm. Once the visual hull has been computed, we formulate the object-specific MRF as described in section 2. Only visual hull based terms are included in the MRF energy construction, and no image based term is used. We estimate the optimal parameters θ_{opt} for the stick-man model by minimizing the MRF energy given in equation (7). Figure 5(i) shows how $\min_{\mathbf{x}} \Psi(\mathbf{x}, \theta)$ changes with different parameters of the stick-man model. It can be clearly seen that the energy surface is locally uni-modal. We use the Powell minimization [6] algorithm for optimization. The graph constructed for the energy minimization procedure has a million nodes connected in a 6 neighbourhood. The time taken by the algorithm of [2] to compute the st-mincut in this graph is 0.3 seconds. In contrast, the dynamic graph cut algorithm only takes 0.01 seconds. For each frame of the video sequence, the Powell minimizer needs roughly 500 function evaluations of $\min_{\mathbf{x}} \Psi(\mathbf{x}, \theta)$ to obtain the solution for $\theta_{\mathbf{opt}}$. Further, as each function evaluation takes roughly 0.15 seconds, we are able to get the pose and reconstruction results in a minute.

Results. Our method is able to obtain accurate object reconstruction results. Additionally, we also obtain an accurate estimate of the pose parameters of the subject. The reconstruction and pose estimation results for a few frames are shown in figure 5(ii).

5 Conclusions

This paper sets out a Bayesian framework for 3D object reconstruction which allows for the integration of 'strong' object-specific and 'weak' smoothness priors with a data based likelihood term. We showed how simple deformable and articulated models can be used as strong priors to overcome the ambiguity plaguing the reconstruction problem. The results of our experiments show that this formulation is not only able to obtain good reconstruction results from noisy data, but also provides us with an accurate estimate of the state of the object, which is quite useful in applications such as human pose inference.

References

1. S. Bhatia, L. Sigal, M. Isard, and M.J. Black. 3d human limb detection using space carving and multi-view eigen models. In *ANM Workshop*, volume I, page 17, 2004.
2. Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *PAMI*, 26(9):1124–1137, September 2004.
3. Y.Y. Boykov and M.P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. In *ICCV*, pages 105–112, 2001.
4. M. Bray, P. Kohli, and P.H.S. Torr. Posecut: Simulataneous segmentation and 3d pose estimation of humans using dynamic graph cuts. In *ECCV*, pages 642–655, 2006.
5. Meijster et al. A general algorithm for computing distance transforms in linear time. *MMAIS-Processing*, pages 331–340, 2000.
6. Press et al. *Numerical recipes in C*. Cambridge Uni. Press, 1988.
7. P. Kohli and P. Torr. Efficiently solving dynamic markov random fields using graph cuts. In *ICCV*, 2005.
8. V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In *ECCV*, volume III, page 82 ff., 2002.

9. V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? In *ECCV*, volume III, page 65 ff., 2002.
10. M.P. Kumar, P.H.S. Torr, and A. Zisserman. Obj cut. In *CVPR*, volume I, pages 18–25, 2005.
11. K.N. Kutulakos and M. Seitz. A theory of shape by space carving. *IJCV*, 38(3), 2000.
12. D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47(1-3):7–42, 2002.
13. D. Snow, P. Viola, and R. Zabih. Exact voxel occupancy with graph cuts. In *CVPR*, 2000.
14. R. Szeliski. Rapid octree construction from image sequences. *CVGIP*, 58:23–32, 1993.
15. G. Vogiatzis, P.H.S. Torr, and R. Cipolla. Multi-view stereo via volumetric graph-cuts. In *CVPR*, volume II, pages 391–398, 2005.