

Refining implicit function representations of 3-D scenes

Matthew Grum and Adrian G. Bors

Dept. of Computer Science, University of York, York YO10 5DD, UK
{grum, adrian.bors}@cs.york.ac.uk

Abstract

This paper considers the problem of modelling a 3-D scene from calibrated images taken from multiple viewpoints. The initial 3-D information is acquired using probabilistic space carving which provides a voxel representation consistent with the given set of images. The scene is afterwards modelled as an implicit surface using radial basis functions (RBF). The mixture of multiorder basis functions models a smoothed 3-D scene representation while providing compactness. We use correspondences between pairs of image patches in order to update the RBF centres for improving the 3-D scene representation. The RBF centre updating leads to improving the consistency between the 3-D model and the given set of images. The proposed method is applied on a complex 3-D scene displaying various objects.

1 Introduction

Three dimensional object reconstruction from several images has lately attracted considerable research interest [1, 8, 9, 12]. Nevertheless, real scenes are very complex and involve several objects, usually occluding each other, while the effects of illumination and material reflectivity cannot be ignored. The aim of this study is to reconstruct the entire 3-D scene from a sparse set of images by estimating both shape and texture.

Space carving is a method which assigns voxels to a 3-D object or to its background using the photoconsistency of a specific point with all its corresponding pixels from the given set of images [9, 10, 12]. There is a lot of uncertainty in the evaluation of the probabilities required for space carving, caused by the presence of occlusions, surface discontinuities, variation in the illumination conditions, camera calibration errors, etc. The resulting voxel model from space carving is invariably noisy and often contains disconnected components. Holes and excessively enlarged 3-D features emerge in the resulting voxel model [9]. Surface refinement for mesh models initialised from volumetric reconstruction has been performed in [5, 6]. In this paper we propose to employ a radial basis function (RBF) in order to model the surface of the space carved data. RBF methods are known for their data fitting, interpolation and generalisation properties and have been widely used in pattern recognition. In our case we want to represent a smooth surface which interpolates the voxels from the surface as accurately as possible to the real scene. Moreover, the RBF model would require only few parameters when compared to the voxel model in order to represent the scene. Implicit RBFs have been shown to represent well surfaces in [4, 11].

In this paper we use the multiorder basis function, proposed by Chen and Suter, which fulfils a smoothness constraint in the first, second and third order Laplacian [2]. The surface of the objects from the scene is calculated as the zero level set of a weighted mixture of basis functions. The basis functions centres are randomly initialised by using a Poisson sphere random sampling scheme [3, 4].

Certain errors are propagated from the voxel model to the implicit surface resulting in surface variations that do not correspond to the actual scene. In this paper we propose to correct such errors by improving the consistency of the 3-D model with the given set of images. In order to achieve good reconstruction accuracy we need to select a wide baseline pair of images with good texture. The pair of images contain the projection of the same part of the 3-D scene, defined around radial basis function centres. An updating formula is derived such that the centre of a certain RBF unit is modified in order to fulfil the consistency between the two projections of the 3-D scene. The proposed methodology is applied on a complex scene representing several objects. The modelling of a 3-D scene using space carving and the modelling using RBF is described in Section 2. The initialisation of the RBF parameters as well as their subsequent updating is described in Section 3. Experimental results are provided in Section 4, while the conclusions of this study are drawn in Section 5.

2 Model initialisation

2.1 Space carving

Let us assume that we have N images of a scene $\{\mathbf{I}_j | j = 1, \dots, N\}$, acquired from various viewpoints by calibrated cameras whose projective matrices \mathbf{P}_j with respect to the scene have been properly calculated. We would like to reconstruct the 3-D scene represented by geometry as well as colour (texture) information. One of the most popular approaches for representing 3-D scenes from multiple images is the space carving algorithm [1, 10, 12]. Probabilistic space carving starts with a parallelepiped formed from voxels. At each iteration, a set of voxels is selected and their consistency with the given set of images is verified. Two assumptions are tested: if a voxel is part of the scene $\mathbf{x} \in \mathcal{V}$, and if it is not, $\mathbf{x} \notin \mathcal{V}$, where \mathbf{x} represents a voxel and \mathcal{V} is the volumetric scene to be estimated. The evaluation of the probability in each image \mathbf{I}_j takes into account its corresponding projection matrix \mathbf{P}_j and checks the photoconsistency of a voxel with its corresponding pixels. Usually, uncertainty arises in the evaluation of the probabilities associating voxels with corresponding pixels from images. Consequently, the resulting volumetric model is invariably noisy. Esteban and Schmitt [5] proposed to use the visual hull in order to initialise a surface mesh which can be deformed under the influence of photoconsistency constraints. In the following we propose to use implicit function modelling estimated from the 3-D voxel data provided by the space carving algorithm.

2.2 Implicit surfaces using radial basis functions

Radial basis functions (RBF) are known for their data fitting, interpolation and generalisation properties [4, 11]. In our approach we use the voxel representation provided by the space carving algorithm by properly interpolating the voxels and smoothing the surfaces in the scene. Moreover, an RBF model would require fewer parameters in order to

represent the scene. The surface of the 3-D scene is modelled as a zero level set of a function, $f(\mathbf{z}) \geq 0$. In our approach, $f(\mathbf{z})$ is an RBF mixture consisting of M basis functions calculated at location \mathbf{z} as :

$$f(\mathbf{z}) = \sum_{i=1}^M w_i \phi(\|\mathbf{z} - \mu_i\|) + u(\mathbf{z}) \quad (1)$$

where $\phi(\cdot)$ is the basis function, considered radially symmetric, $\|\cdot\|$ is the Euclidean distance, μ_i is the basis function centre, and $u(\mathbf{z})$ is a polynomial component. The function $f(\mathbf{z})$ is defined as positive inside the 3-D volume and negative outside. For $f(\mathbf{z}) = 0$ we obtain the surface enveloping the 3-D voxel model.

Gaussian RBF functions which are widely used in pattern recognition have been found to oversmooth [4]. Chen and Suter derived a basis function which fulfils a constraint in the first, second and third order Laplacian, [2] :

$$-\delta\Delta f(\mathbf{z}) + \Delta_2 f(\mathbf{z}) - \tau\Delta^3 f(\mathbf{z}) = 0 \quad (2)$$

where Δ is the Laplacian operator in 3-D, δ is a parameter controlling the first order smoothness and τ controls the third order smoothness. The function that minimises the energy function from (2) is called multiorder basis function [2, 4] :

$$\phi(\|\mathbf{z} - \mu_i\|) = \frac{1}{4\pi\delta^2\|\mathbf{z} - \mu_i\|} \left(1 + \frac{\beta e^{-\|\mathbf{z} - \mu_i\|\sqrt{\alpha}}}{\alpha - \beta} - \frac{\alpha e^{-\|\mathbf{z} - \mu_i\|\sqrt{\beta}}}{\alpha - \beta} \right) \quad (3)$$

where

$$\alpha = \frac{1 + \sqrt{1 - 4\tau^2\delta^2}}{2\tau^2} ; \beta = \frac{1 - \sqrt{1 - 4\tau^2\delta^2}}{2\tau^2} \quad (4)$$

are parameters which describe the shape of the basis function.

3 RBF parameter calculation

3.1 Initialising the RBF parameters

The RBF function has the property to approximate well the data in a specific neighbourhood as shown by the expression (3). The RBF function from (3) has the maximum in the centre μ_i and quickly falls toward zero when the distance from its centre location increases. In this study we use the Poisson sphere random sampling scheme for initialising the RBF centres [4]. This algorithm has been proposed in [3] by Cook for solving the aliasing problem in computer graphics. A Poisson sphere distribution is a 3-D random point distribution in which all sphere centres are approximately equally distributed in space. Let us consider a set of spheres as $\mathbf{S}(\mu_k, \rho)$, $k = 1, \dots, M$, each centred at μ_i and with identical radius, ρ . The sphere radius ρ depends on the size of the voxel model, $|\mathcal{V}|$. The number of basis functions M and consequently that of spheres depends on the desired level of surface approximation and smoothness.

Centres of spheres are randomly generated within the given voxel space such that they fulfil the following conditions :

$$\|\mu_i - \mu_j\| \geq 2\rho \quad (5)$$

where 2ρ is the minimum distance between two sphere centres i and j . Each sphere determines a partition in the voxel model depending on the local compactness. Let us

consider a set of at least T connected voxels which are located within a radius of ρ from the centre of the sphere :

$$\{\mathbf{x}_c \in \mathbf{S}(\mu_i, \rho) | \mathbf{x}_c \in \mathcal{V}, \|\mathbf{x} - \mu_i\| < \rho, |\mathbf{x}_c| \geq T\} \quad (6)$$

where $|\cdot|$ denotes set cardinality. The spheres which contain very few voxels as well as unconnected voxels are discarded. The sphere generating algorithm terminates when the voxel model is completely covered with spheres. Let us assume that a total of M valid spheres $\mathbf{S}(\mu_i, \rho)$ are generated, each associated with an RBF centre, μ_i . The parameters τ and δ determine the smoothing of the resulting implicit surface. These parameters are chosen depending on the chosen resolution, the size of the voxel model $|\mathcal{V}|$, and on the desired level of smoothing [4].

We form the following system of equations :

$$\begin{bmatrix} \phi(r_{11}) + \lambda_1 & \dots & \phi(r_{1M}) & 1 \\ \vdots & \ddots & \vdots & \mathbf{1} \\ \phi(r_{M1}) & \dots & \phi(r_{MM}) + \lambda_M & 1 \\ 1 & \mathbf{1} & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_M \\ u_0 \end{bmatrix} = \begin{bmatrix} f(\mu_1) \\ \vdots \\ f(\mu_M) \\ 0 \end{bmatrix} \quad (7)$$

where $r_{ij} = \|\mu_i - \mu_j\|$ is the Euclidean distance between two centres, λ_i for $i = 1, \dots, M$ are added to the diagonal elements in order to condition better the matrix as in [4] and $u(\mathbf{z}) = u_0$. For calculating the weighting factors w_i , $i = 1, \dots, M$ we evaluate the basis functions $\phi(\cdot)$ for the distances between pairs of centres r_{ij} . We consider $f(\mu_i) = 0$ for imposing the condition that most basis functions are located on the separation surface. Certain centres correspond to the control basis functions, *i.e.* which have their centres either inside the model or outside it. The weights w_i , $i = 1, \dots, M$ are calculated by inverting the matrix associated with basis function centres. Given the proposed RBF centre initialisation described in the previous Section, the matrix from (7) is non-singular and consequently invertible.

3.2 Updating the RBF centres using image disparity

The previous approaches adopted in space carving have been restricted to considering per voxel consistency. In this Section we describe how the accuracy of the surface can be improved by considering the image consistency across larger areas of the surface. Invariably, given various sources of errors, the surface described by the implicit function $f(\mathbf{z})$ may not fit with its corresponding areas of the images. In this Section we describe how to find an updating transformation applied on the basis function parameters in order to improve the consistency between the 3-D model given by $f(\mathbf{z})$ and the image set $\{\mathbf{I}_j | j = 1, \dots, N\}$. The surface, as defined by (7), passes through the radial basis function centres. We can control the surface by changing the locations of the RBF centres. The first order approximation of an RBF consists of the plane tangent to its surface in the neighbourhood of its centre. For a small area well defined around the RBF centre we assume that the surface function $f(\mathbf{z})$ can be locally approximated by a planar patch. Let us assume that there are two images which contain projections of the 3-D patch. A plane in 3-D, such as the one which approximates locally the surface around the basis function centre, induces a homography \mathbf{H} between pairs of images [8]. By calculating \mathbf{H} from pairs of images it is possible to recover the parameters of this plane and correct the basis function centre and thus that part of the surface, by constraining it to lie on that plane.

A surface patch, corresponding to an RBF centre, is selected if it displays a sufficient amount of detail which can be used for finding matches between pairs of images. For each chosen patch we select a pair of images such that they provide the smallest angle between their positions and the surface patch normal. The angle between the camera locations and the surface patch should be as large as possible in order to provide an appropriate baseline to recover positions. Let \mathbf{P} and \mathbf{P}' be two 3×4 matrices which describe the camera projection from 3-D coordinates to homogenous image coordinates for the selected pairs of images representing the patch. Let $\mathbf{y} = [u, v, 1]^T$ be the projection of a point in the patch from the first camera and $\mathbf{y}' = [u', v', 1]^T$ be the corresponding point from the second camera. These points are related by $\mathbf{y}' = \mathbf{H}\mathbf{y}$. Let us assume that the selected patch belongs to a plane ψ , where $\mathbf{z}^T \psi = 0$ for all the points \mathbf{z} which lie on ψ . The homography \mathbf{H} between the pair of images is given as, [8] :

$$\mathbf{H} = \mathbf{A} - \mathbf{a}\mathbf{v}^T \quad (8)$$

where \mathbf{A} and \mathbf{a} are a 3×3 matrix and a 3×1 vector, respectively, given by :

$$[\mathbf{A} \mid \mathbf{a}] = \mathbf{P}' \left[\begin{array}{c|c} \mathbf{P} & \\ \hline 0 & 0 \ 0 \ 1 \end{array} \right]^{-1} \quad (9)$$

and \mathbf{v} is a 3×1 vector, representing the displacement between the two images, given by the following expression :

$$\begin{bmatrix} \mathbf{v} \\ 1 \end{bmatrix} = \left(\left[\begin{array}{c|c} \mathbf{P} & \\ \hline 0 & 0 \ 0 \ 1 \end{array} \right]^{-1} \right)^T \psi \quad (10)$$

Given a point in one image, the corresponding point in another image can be constrained to lie on a line known as the epipolar line [8]. Epipolar lines depend only on the imaging geometry and not on the shape of the scene, so it is possible to transform the images, using \mathbf{v} , in order to correspond to a pair of rotated ‘virtual cameras’, whose epipolar lines are all horizontal and co-linear. This process is known as rectification and is often performed as an initial step in stereo algorithms [7].

Let \mathbf{R} and \mathbf{R}' be the rectifying 3×3 matrix transformations. The rectified images of the patch are now related by considering a matrix \mathbf{H}_R :

$$\mathbf{R}'\mathbf{y}' = \mathbf{H}_R\mathbf{R}\mathbf{y} \quad (11)$$

The homography \mathbf{H} can be calculated by taking into account the rectifying transformations :

$$\mathbf{H} = \mathbf{R}'^{-1}\mathbf{H}_R\mathbf{R} \quad (12)$$

After the rectification, the epipoles are horizontal, and \mathbf{H}_R is guaranteed to map each v -coordinate to its corresponding value in each pair of images. Consequently, it can be expressed as :

$$\mathbf{H}_R = \begin{bmatrix} s & k & t \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (13)$$

where s , k and t , correspond to scaling, skew and translation, respectively (all in the u direction). To calculate these parameters, the images of the patch are divided into l rows

of pixels. When considering a single row of pixels, the skew and translation act together to produce a single horizontal offset, o , since the v coordinate of each pixel is the same.

The normalised cross-correlation is computed between each pair of rows at different scale and offset values. The values which result in the lowest score, corresponding to the best match, are recorded. As the scale should be the same for all rows, s is taken to be the median of the values found for each row. Any values significantly outside the median are deemed to be unreliable and are discarded. Using the offsets from all rows, the skew and translation parameters k and t can be calculated by solving a linear system:

$$\begin{bmatrix} k \\ t \end{bmatrix} = \begin{bmatrix} v_1 & 1 \\ \vdots & \vdots \\ v_l & 1 \end{bmatrix}^{-1} \begin{bmatrix} o_1 \\ \vdots \\ o_l \end{bmatrix} \quad (14)$$

where v_l is the v coordinate of row l .

With \mathbf{H} at hand from (12), we calculate the displacement vector \mathbf{v} between the pair of images corresponding to the given patch from (8). Consequently, the location of the plane ψ which should contain the basis function is calculated using (10). The location of the basis function centre is updated as :

$$\mu'_i = \mu_i + \mathbf{n} \frac{\mu_i^T \psi}{\mathbf{n}^T \mathbf{n}} \quad (15)$$

where \mathbf{n} is the surface normal direction of the plane ψ , and the i th basis function centre μ_i is updated to μ'_i , while being constrained to lie on the plane ψ .

The matching based on cross-correlation requires that the colour variance in the image patch is above a certain threshold in order to find the offsets uniquely. Basis functions corresponding to patches which do not fulfil this condition are not updated by this procedure. Additionally, false matches may be obtained due to the image noise or to patches which span the physical boundary of an object. A limit is placed on the maximum distance that a centre can move in order to prevent this from causing further errors in the surface. Some of the basis function centres will converge towards neighbouring locations on the 3-D surface causing singularity in the matrix from equation (7). If centres of multiple basis functions occur in the immediate proximity of each other after updating, only one will be preserved while the others will be removed.

4 Experimental results

The method outlined in this paper was tested on a real scene comprised of multiple objects. For the experiments, 12 images of the scene were captured from various viewpoints. A selection of four images is shown in Fig. 1. As it can be observed from this Figure, the objects exhibit various shapes and surface properties and occlude each other in different views. Voxel carving assumes the camera positions (extrinsic calibration) to be known *a priori*. Targets printed on rectangular boards were placed around the outside of the scene in order to provide the necessary information for camera calibration.

The initial voxel model was provided by the probabilistic space carving algorithm [1]. This algorithm assumes the scene to be contained within a finite bounding volume. In this case, the background was manually segmented. The resulting model, which contains 773660 voxels, is shown in Fig. 4(a) and Fig. 4(b) for two different viewpoints. Considering the voxel representation, 4440 radial basis functions were sampled and used to



Figure 1: Four images of a complex scene taken from various viewpoints.

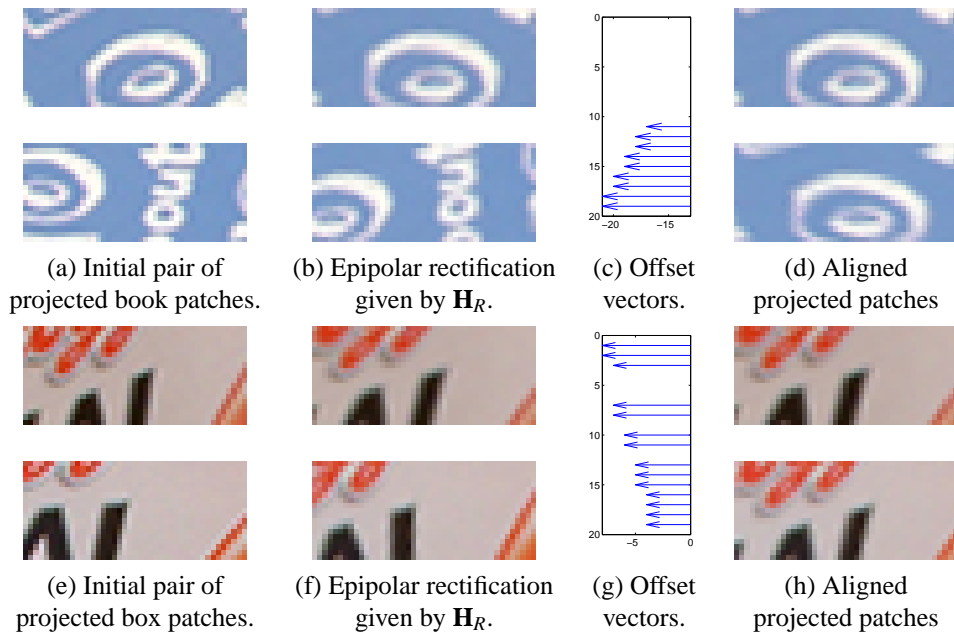
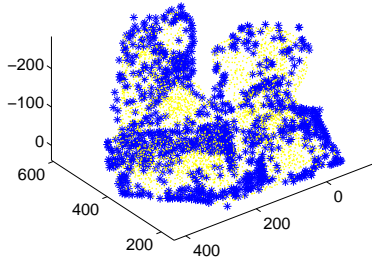
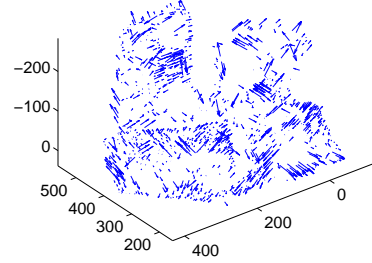


Figure 2: Two pairs of patches and their correction.



(a) RBF centres after updating.
Updated centres are marked by "*".



(b) RBF centres correction vectors.

Figure 3: Directions of centre updating and their corrected positions.

fit the implicit surface as shown in Figs. 4(c) and 4(d). A total of 1408 basis functions met the criteria for updating (the visibility and the presence of sufficient local variation as given by the colour variance). Of these, a suitable match was obtained in 1075 cases. The smoothing parameters, considered identical for all RBFs, are $\delta = 25$, $\tau = 0.01$, while the centres are scaled such that they fit in a cube of size $1 \times 1 \times 1$. The functions which were successfully updated are shown in Fig. 3(a), marked with stars, while all the other basis functions are marked with dots.

Two pairs of patches from the raw images, which are the projections of two different 3-D scene regions, one corresponding to the book and another to the box, are shown in Figs. 2(a) and 2(e), respectively. The images from each pair are related by means of \mathbf{H} , according to (8). The epipolar correction, as given by \mathbf{H}_R from (13), is shown in Figs. 2(b) and 2(f), the offset vectors calculated from equation (14) are provided in Figs. 2(c) and 2(g), and the aligned patches after applying the transformation \mathbf{H} to the second image of each pair is illustrated in Figs. 2(d) and 2(h). Vectors representing the movement of centres in 3-D to the corrected positions, according to (15), are shown in Fig. 3(b). The updated surface, after correcting the RBF centres, for the two viewpoints, is shown in Fig. 4(e) and Fig. 4(f), respectively. The surface of the horizontal book and vertical box is clearly improved. However, the shape of certain objects, the kettle in particular, is not well modelled due to their irregular shapes, lack of texture and surface specularities.

For numerical assessment we check the consistency between the surface of the book from the estimated 3-D model with that from the real scene. The surface of the book in the centre of the scene is planar and we measure the deviation from the planarity in the estimated 3-D model. This deviation, measured in millimetres, was estimated for the voxel model, the initial RBF surface, calculated according to the description from Section 3.1 as well as for the surface updated according to the algorithm provided in Section 3.2. The mean deviation was found to be 11.59 mm for the voxel model, 3.63 mm for the initial RBF estimation and 1.04 mm for the updated model. These numerical results together with the visual interpretation from Fig. 4 prove the capabilities of the proposed algorithm to improve the surface representation when considering the proposed RBF centre updating method based on image disparity estimation.

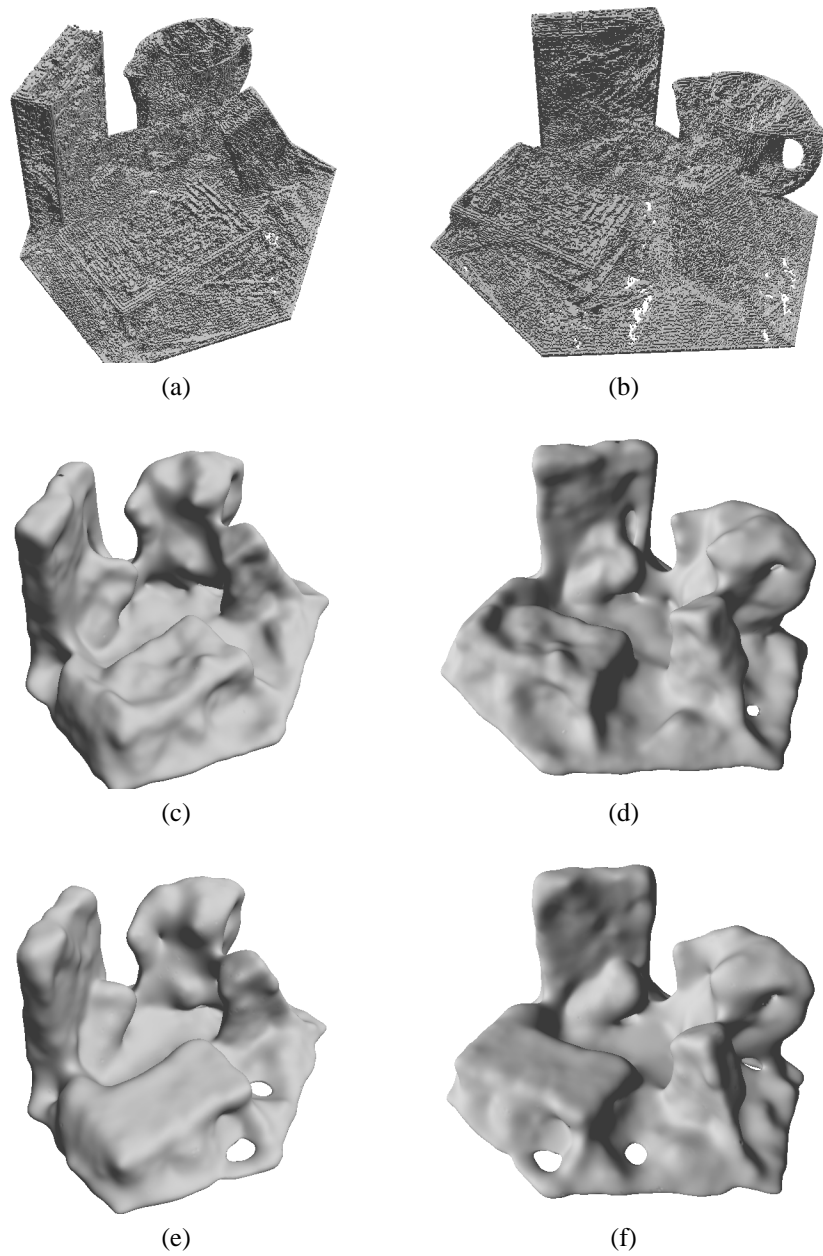


Figure 4: Voxel representation and RBF surface modelling of the scene for two different view angles. (a), (b) voxel representation; (c), (d) initial RBF model; (e), (f) updated RBF model.

5 Conclusion

A complex 3-D scene surface modelling method using multiple images, taken from various viewpoints, is proposed in this paper. A voxel representation is estimated using the space carving algorithm. Implicit multiorder radial basis functions are employed in order to model the separation surface between the voxel model and the exterior. The RBF model produce a smoother 3-D scene than the voxel representation while requiring much less parameters. The 3-D representation is improved by using an RBF centre updating algorithm. The proposed algorithm estimates the disparity errors between pairs of images after recovering their perspective distortions. The resulting 3-D surface representation can be easily rendered and manipulated by geometrical transformations.

References

- [1] A. Broadhurst, T. W. Drummond, and R. Cipolla. A probabilistic framework for space carving. In *Proc. ICCV, vol. 1*, pages 388–393, Vancouver, BC, Canada, 2001.
- [2] F. Chen and D. Suter. Multiple order Laplacian spline - including splines with tension. Technical Report MECSE 1996-5, Monash University, Australia, July 1996.
- [3] R. L. Cook. Stochastic sampling in computer graphics. *ACM Transactions on Graphics*, 5(1):51–72, 1986.
- [4] H. Q. Dinh, G. Turk, and G. Slabaugh. Reconstructing surfaces by volumetric regularization using radial basis functions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(10):1358–1371, 2002.
- [5] C. H. Esteban and F. Schmitt. Silhouette and stereo fusion for 3D object modeling. *Computer Vision and Image Understanding*, 96(3):367–392, 2004.
- [6] Y. Furukawa and J. Ponce. Carved visual hulls for image-based modeling. In *Proc. ECCV, part I, LNCS 3951*, pages 564–577, Graz, Austria, 2006.
- [7] A. Fusiello, E. Trucco, and A. Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22, 2000.
- [8] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition, 2004.
- [9] H. Kim and I. S. Kweon. Appearance-cloning: photo-consistent scene recovery from multi-view images. *Int. Journal of Computer Vision*, 66(2):163–192, 2006.
- [10] K. Kutulakos and S. M. Seitz. A theory of shape by space carving. *Int. Journal of Computer Vision*, 38(3):198–218, 2000.
- [11] B. Scholkopf, J. Giesen, and S. Spalinger. Kernel methods for implicit surface modeling. In *Proc. of Neural Information Proc. Systems (NIPS)*, pages 1193–1200, 2004.
- [12] G. G. Slabaugh, W. B. Culbertson, T. Malzbender, M. R. Stevens, and R. W. Schafer. Methods for volumetric reconstruction of visual scenes. *Int. Journal of Computer Vision*, 57(3):179–199, 2004.