Pattern Recognition 47 (2014) 326–343

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

3D modeling of multiple-object scenes from sets of images

Matthew Grum, Adrian G. Bors *

Dept. of Computer Science, University of York, York YO10 5GH, UK

ARTICLE INFO

Article history:
Received 14 July 2011
Received in revised form
19 March 2013
Accepted 30 April 2013
Available online 23 May 2013

Keywords:
3D scene modeling
Disparity correction
Shape-from-contours

ABSTRACT

This paper proposes a new approach for multi-object 3D scene modeling. Scenes with multiple objects are characterized by object occlusions under several views, complex illumination conditions due to multiple reflections and shadows, as well as a variety of object shapes and surface properties. These factors raise huge challenges when attempting to model real 3D multi-object scene by using existing approaches which are designed mainly for single object modeling. The proposed method relies on the initialization provided by a rough 3D model of the scene estimated from the given set of multi-view images. The contributions described in this paper consists of two new methods for identifying and correcting errors in the reconstructed 3D scene. The first approach corrects the location of 3D patches from the scene after detecting the disparity between pairs of their projections into images. The second approach is called shape-from-contours and identifies discrepancies between projections of 3D objects and their corresponding contours, segmented from images. Both unsupervised and supervised segmentations are used to define the contours of objects.

1. Introduction

Single 3D object reconstruction from multiple images has lately attracted considerable research interest. However, real scenes contain multiple objects where occlusions can occur under several views. Moreover, object shadows and complex surface shading variation happen due to multiple light reflections. There are various approaches for modeling 3D shapes using image photo-consistency, shape from shading, from silhouettes, or from texture. Multi-camera stereo vision has been addressed as a geometry correspondence problem in [1]. Space carving is a method which assigns voxels to a 3D object, or carves them away from its background using the photo-consistency of a specific point with all its corresponding pixels from the given set of images [2–5]. The voxel model produced by space carving is invariably noisy and often contains disconnected components [4]. A semi-supervised approach was used for space carving in [6]. At a higher scene representation level, patches have been used for modeling the surface of 3D objects using multi-view constraints [7–9]. Combining multi-view stereo and surface radiance estimation was employed for modeling 3D objects with non-Lambertian properties in [10]. The volumetric graph-cut method employs the visual hull of the scene as a constraint on its topology and for inferring occlusions [11]. Surface refinement for mesh models initialized from volumetric reconstructions was used in [12,13]. Radial Basis Functions (RBF) are known for their data fitting, interpolation and generalization properties and have been widely used in pattern recognition [14]. Implicit RBFs have been used for modeling object surfaces in [15].

In shape-from-silhouettes, a set of silhouettes extracted from images is used to model the 3D scene by generating the convex hull produced by a union of projection cones [16,17]. Each of these cones has its apex located at one of the camera positions whilst its axis orientation is given by the corresponding camera parameters. The visual hull of the scene results from interpolating the data provided by the 3D scene silhouettes by using various assumptions about the local surface smoothness as in [18,19]. Similarly to the 3D Hough transform from [14], shape reconstruction was performed using the duality between the 3D shape and its corresponding set of tangent planes to its surface calculated using the edges extracted from images [20]. These methods depend on the 3D interpolation in the regions for which silhouettes or other edges are not available. In [12] an energy function with two components was used for fusing texture and silhouette information for single 3D object reconstruction. The energy function was used to guide a deformable model for representing the scene.

In this paper we propose a multistage methodology for 3D representation of scenes with multiple objects. The 3D scene is initialized using probabilistic space carving [21]. The carved voxels are then used for representing an implicit 3D surface model based on an implicit radial basis functions (RBF) model [22]. The second stage of the proposed methodology corrects the errors in the 3D

* Corresponding author. Tel.: +44 190443325574; fax: +44 1904432767.
E-mail address: adrian.bors@york.ac.uk (A.G. Bors).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.04.020

scene by improving its consistency with the image set content. Two different consistency approaches are used for updating the 3D scene surface such that it matches the information from the given image set: using disparity correction of textured image areas and by enforcing the consistency with the segmented object contours from images. These two different consistency enforcement procedures will aim to correct errors in different parts of the scene: the first one is required in areas with lots of image variation while the second is used in areas characterized by well defined object contours. The first approach uses the disparity in projections of 3D patches in order to correct their miss-alignment and consequently the 3D scene representation. The second updating procedure uses shape-from-contours for correcting the 3D scene representation. The object contours are extracted from images using either unsupervised or supervised segmentation. The 3D surface initialization using space carving and implicit RBFs is described in Section 2. The scene updating method using the disparity between image regions is described in Section 3, while the updating method using object contours is explained in Section 4. The experimental results are presented in Section 5, while the conclusions are drawn in Section 6.

2. Initializing the 3D scene model

Let us assume that we have a set of images $\mathcal{I} = \{\mathbf{I}_i | i = 1, \ldots, n\}$, each taken from a different viewing angle, which are characterized by their projection matrices, defined by the camera parameters, $\mathcal{P} = \{\mathbf{P}_i | i = 1, \ldots, n\}$. 3D scene modeling from images is a problem of recovering missing information from sparse data given a set of constraints. Various approaches have been used for modeling 3D scenes from images, usually representing a single centered object. In the case of scenes with multiple objects, occlusions may occur in several images. Moreover, shadows and complex shading variation happen due to multiple light reflections in the scene. The proposed 3D scene reconstruction methodology is initialized by using probabilistic space carving followed by the implicit radial basis functions representation of the surface.

Space carving produces a voxel model $\mathcal{V}$ of the 3D scene from multiple images $\mathcal{I}$ by using epipolar geometry and photo-consistency, given their camera parameters $\mathcal{P}$. The probabilistic space carving algorithm, proposed in [21], is a variant of the voxel carving approach, in which voxel occupancy is determined by statistical estimation using Bayes theorem. This approach assumes the visibility of each voxel in the scene from several images, which may not hold in multi-object scenes due to occlusions [5]. As a consequence, the photoconsistency measure, used to decide whether a voxel is carved or not, may be compromised and “floating voxels” will result above the surfaces as well as in the gaps between objects [4].

In order to improve the 3D scene surface representation we use implicit radial basis functions (RBF) estimated from the voxel representation. RBFs allow both approximation and data interpolation introducing smoothness and compactness in the 3D scene representation [14,22]. The surface $\mathcal{S}$ is modeled as

$$f(\mathbf{x}) = \sum_{i=1}^{M} w_i \phi(|\mathbf{x} - \mathbf{c}_i|) + p(\mathbf{x}), \tag{1}$$

where we have M radially symmetric basis functions, $\phi(\cdot)$, each characterized by weight w_i and center $\mathbf{c}_i$ while $p(\mathbf{x})$ is a polynomial function, usually considered as a constant. An implicit surface reconstruction is defined by $f(\mathbf{x}) = 0$.

If each radial basis function center $\mathbf{c}_i$ is chosen to coincide with the location of selected observations, then the weights can be calculated by solving a linear system of equations:

$$\begin{bmatrix} \phi(r_{11}) + \lambda_1 & \cdots & \phi(r_{1M}) & 1 \\ \vdots & \ddots & \vdots & \mathbf{1} \\ \phi(r_{M1}) & \cdots & \phi(r_{MM}) + \lambda_M & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_M \\ p_0 \end{bmatrix} = \begin{bmatrix} f(\mathbf{c}_1) \\ \vdots \\ f(\mathbf{c}_M) \\ 0 \end{bmatrix} \tag{2}$$

where $r_{ij} = |\mathbf{c}_i - \mathbf{c}_j|$, and λ_i, $i = 1, \ldots, L$ are terms added to the diagonal in order to allow for $f(\mathbf{x})$ to deviate by a small margin from the actual observations, as discussed in [23]. The observations are points which lie on the 3D surface imposing constraints. In addition to these, a small number of external or internal constraints must be specified, with negative or positive values, in order to provide orientation to the surface and prevent the trivial null solution for $f(\mathbf{x}) = 0$. The coordinates of voxels were used as surface constraints in [15] as well. It is not feasible to use all the voxels from $\mathcal{V}$ and a regular sub-sampling of the discrete volume cannot be used because it can lead to singularity when attempting to solve (2). Consequently, we propose to use a Poisson sphere random sampling scheme, corresponding to the 3D extension of the Poisson disc from [24], for initializing the RBF centers. Spheres of a given radius are randomly generated in the space of $\mathcal{V}$ such that no two spheres would overlap. Spheres containing very few vertices would be removed. The vector median of the voxels within each sphere [25], is used to determine $\mathbf{c}_i$. For small values of λ_i, $i = 1, \ldots, M$, the corresponding RBFs represent regions of voxels defined by high photoconsistency [22].

Gaussian functions tend to over-smooth surfaces, and in this study we use the multi-order basis function, which was derived in [26] after imposing a combination of first, second and third order smoothness constraints:

$$-\delta \Delta \phi(\mathbf{x}) + \Delta^2 \phi(\mathbf{x}) - \tau \Delta^3 \phi(\mathbf{x}) = 0, \tag{3}$$

where $\phi(\mathbf{x})$ represents the basis function to be derived and the parameters τ and δ define the smoothness constrains. The resulting function is given by

$$\phi(r) = \frac{1}{4\pi\delta^2 r}\left(1 + \frac{\mu e^{-\sqrt{\nu} r}}{\nu - \mu} - \frac{\nu e^{-\sqrt{\mu} r}}{\nu - \mu}\right), \tag{4}$$

where $r = |\mathbf{x} - \mathbf{c}|$, represents the distance from the center $\mathbf{c}$ and the shape parameters ν, μ are given by

$$\nu = \frac{1 + \sqrt{1 - 4\tau^2\delta^2}}{2\tau^2}; \quad \mu = \frac{1 - \sqrt{1 - 4\tau^2\delta^2}}{2\tau^2}. \tag{5}$$

This type of basis function allows more detail to be retained in the resulting 3D reconstruction without sacrificing smoothness, as it was shown in [15].

Besides good smoothness and interpolation properties, RBFs provide a compact representation of the scene surface $\mathcal{S}$, requiring in total $2M + 2$ parameters instead of a huge amount of voxel locations. An RBF representation of the surface can easily be manipulated by affine transformations and locally corrected whenever necessary, as described in the following sections. However, both space carving and RBF modeling are limited in their performance when attempting to represent scenes with multiple objects. In the following we propose two successive updating stages in order to correct the inconsistencies arising in 3D representations of multiple objects from multi-view images.

3. Updating scene surfaces using image disparities

The voxel model obtained by means of the probabilistic space carving often contains many spurious voxels resulting from the uncertainty in the illumination and camera positions. Whilst small scale protrusions and gaps are smoothed or interpolated over by