Available online at www.sciencedirect.com
ScienceDirect
Image and Vision Computing 26 (2008) 397–404
www.elsevier.com/locate/imavis

Reconstructing relief surfaces

George Vogiatzis [a,*], Philip H.S. Torr [b], Steven M. Seitz [c], Roberto Cipolla [a]

[a] Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK
[b] Department of Computing, Oxford Brookes University, Wheatley, Oxford OX33 1HX, UK
[c] Department of Computer Science and Engineering, University of Washington, Seattle, WA 98195-2350, USA

Received 25 February 2005; received in revised form 8 December 2006; accepted 10 January 2007

Abstract

This paper generalizes Markov Random Field (MRF) stereo methods to the generation of surface relief (height) fields rather than disparity or depth maps. This generalization enables the reconstruction of complete object models using the same algorithms that have been previously used to compute depth maps in binocular stereo. In contrast to traditional dense stereo where the parametrization is image based, here we advocate a parametrization by a height field over any base surface. In practice, the base surface is a coarse approximation to the true geometry, e.g., a bounding box, visual hull or triangulation of sparse correspondences, and is assigned or computed using other means. A dense set of sample points is defined on the base surface, each with a fixed normal direction and unknown height value. The estimation of heights for the sample points is achieved by a belief propagation technique. Our method provides a viewpoint independent smoothness constraint, a more compact parametrization and explicit handling of occlusions. We present experimental results on real scenes as well as a quantitative evaluation on an artificial scene.

Keywords: Stereo; Multi-view; Reconstruction; MRF; Belief propagation; Volumetric; Vision

1. Introduction

Inferring the dense 3D geometry of a scene from a set of photographic images is a computer vision problem that has been extensively studied. Work in this area can be roughly divided into two classes: (1) techniques for computing depth maps (image-based parameterization), and (2) volumetric methods for computing more complete object models.

In the first class, *image based parameterization* of shape, a reference image is selected and a disparity or depth value is assigned to each of its pixels using a combination of image correlation and regularization. Scharstein and Szeliski provide an excellent review for image based methods [21]. These problems are often formulated as minimisations of Markov Random Field (MRF) energy functions providing a clean and computationally-tractable formulation, for which good approximate solutions exist using Graph cuts [2,15,20,11] or Loopy Belief Propagation [24]. They can also be formulated as continuous PDE evolutions on the depth maps [23]. However, a key limitation of these solutions is that they can only represent depth maps with a unique disparity per pixel, i.e. depth is a function of image point. Capturing complete objects in this manner requires further processing to merge mul tiple depth maps [18], a complicated and error-prone procedure. A second limitation is that the smoothness term imposed by the MRF is viewpoint dependent, in that if a different view was chosen as the reference image the results could be quite different.

The second class of techniques uses a *volumetric parameterization* of shape. In this class are well-known techniques like Space Carving [16] and level-set stereo [6]. There are also hybrid approaches that optimize a continuous functional via a discrete quantisation [19]. While these methods are known to produce high quality reconstructions, run-

* Corresponding author. Tel.: +44 7900023260.
E-mail addresses: gv215@eng.cam.ac.uk (G. Vogiatzis), philiptorr@brookes.ac.uk (P.H.S. Torr), seitz@cs.washington.edu (S.M. Seitz), cipolla@eng.cam.ac.uk (R. Cipolla).

0262-8856/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2007.01.006

ning on high resolution 3D grids is very computationally and memory intensive. Furthermore their convergence properties in the presence of noise are not well understood, in comparison with MRF techniques, for which strong convergence results are known. For Space Carving in particular, there is also no simple way to impose surface smoothness constraints.

In principle MRF stereo methods could be extended to multiple views. The problem is that reasoning about occlusions within the MRF framework is not straightforward because of global interactions between points in space (see [15] for an insightful but costly solution for the case of multi-view depth-map reconstruction). In this paper, we propose extending MRF techniques to the multi-view stereo domain by recovering a general *relief surface*, instead of a depth map. We assume that a coarse *base surface* is given as input. In practice this can be obtained by hand, by shape-from-silhouette techniques or triangulating sparse image correspondences. On this base surface sample points are uniformly and densely defined, and a belief propagation algorithm is used to obtain the optimal height above each sample point through which the relief surface passes. The benefits of our approach are as follows:

(1) General surfaces and objects can be fully represented and computed as a single relief surface.
(2) Optimisation is computationally tractable, using existing MRF solvers.
(3) Occlusions are approximately modelled.
(4) The representation and smoothness constraint is image and viewpoint independent.

1.1. Related work

Our work is inspired by displaced surface modelling methods in the computer graphics community, in particular the recent work of Lee et al. [17], who define a displacement map over subdivision surfaces, and describe a technique for computing such a representation from an input mesh. An advantage of this and similar techniques is that they enable the representation of finely detailed geometry using a simple base mesh.

We also build on work in the vision community on *plane-plus-parallax* [3], *model-based stereo* [5], and *sprites with depth* [22]. All of these techniques provide means for representing planes in the scene with associated height fields. Our work can be interpreted as a generalization of plane-plus-parallax to a surface-plus-height formulation.

Previous mesh-based multi-view stereo techniques operate by iteratively evolving an initial mesh until it best fits a set of images [14,26], or depth maps [10]. Representing finely detailed geometry is difficult for such methods due to the need to manage large and complex meshes. In contrast we assume a fixed base surface and solve only for a height field providing a much simpler way of representing surface detail. We also use a more stable estimation problem with good convergence properties. Ultimately, a hybrid approach that combines surface evolution and height field estimation could offer the best of both worlds and is an interesting topic of future work.

2. Model

The theory of Markov random fields yields an efficient and powerful framework for specifying complex spatial interactions between a number of discrete random variables $h_1, \ldots, h_M$, usually called *sites*. Each site can take one of a number of values or *labels* $H_1, \ldots, H_L$. The first ingredient of the model is a labelling cost function $C_k(h_k)$ that measures how much a site is in agreement with being assigned a particular label. The second ingredient is the interaction between sites, which, in a pairwise MRF such as the one considered in this paper, is modelled through a symmetric neighbourhood relation $\mathcal{N}$ as well as a compatibility cost term $C_{kl}(h_k, h_l)$ defined over neighbouring sites. This cost term measures how compatible the assignment of any two neighbouring labels is. The cost of cliques (fully connected subgraphs) with more than two nodes is set to zero. With these energy functions defined, the joint probability of the MRF is:

$$Pr(h_1, \ldots, h_M) = \frac{1}{Z} \exp\left(-\sum_{k=1}^{M} C_k(h_k) - \sum_{(k,l) \in \mathcal{N}} C_{kl}(h_k, h_l) \right) \tag{1}$$

where Z is a constant.

To bring multi-view stereo into this framework a set of 3D sample points $\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_M$ is defined on a *base surface*. The neighbourhood relation $\mathcal{N}$ defined between the sample points can be obtained in a number of ways, some of which are discussed in the next subsection. At each sample point $\mathbf{X}_k$, the unit normal to the base surface at that point, $\mathbf{n}_k$ is computed. The sites of the MRF correspond to height values $h_1, \ldots, h_M$ measured from the sample points $\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_M$ along the normals $\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_M$ (see Fig. 1 left). The labels $H_1, \ldots, H_L$ are a set of possible height values that variables h_k can take. If the kth site is assigned label h_k then the *relief surface* passes through 3D point $\mathbf{X}_k + h_k\mathbf{n}_k$. To deal with the problem of occlusion, the base surface has to *contain* the relief surface for reasons that will be explained in Section 2.2. Hence if the positive normal direction is defined to be towards the interior of the volume, only positive (inward) heights need be considered. The labelling cost is related to the photo-consistency [16] of the 3D point $\mathbf{X}_k + h_k\mathbf{n}_k$ while the compatibility cost forces neighboring sites to be labelled with 'compatible' heights. The following sections discuss how to define the sample point neighbourhood relation, as well as the two terms of the cost functional in more detail.

2.1. Sample point neighbourhood

The neighbourhood relation between sample points can in principle be obtained by a simple thresholding of the