PIECEWISE-PLANAR 3D RECONSTRUCTION IN RATE- DISTORTION SENSE[°]

Evren İmre, Uğur Güdükbay[+] and A. Aydın Alatan

Department of Electrical & Electronics Engineering, METU
Balgat 06531 Ankara, TURKEY
[+]Department of Computer Engineering, Bilkent University
Bilkent 06533 Ankara, TURKEY
E-Mail: {e105682@, alatan@eee.}metu.edu.tr, gudukbay@cs.bilkent.edu.tr

ABSTRACT

In this paper, a novel rate-distortion optimization inspired 3D piecewise-planar reconstruction algorithm is proposed. The algorithm refines a coarse 3D triangular mesh, by inserting vertices in a way to minimize the intensity difference between an image and its prediction. The preliminary experiments on synthetic and real data indicate the validity of the proposed approach.

Index Terms— Rate-distortion optimal 3D reconstruction, piecewise planar 3D reconstruction

1. INTRODUCTION

In 3D TV applications, dense 3D scene representations have a considerable potential to remove visual redundancy in multi-view video and to improve the overall compression rate of the 3D TV bit-stream. However, in order to realize this goal, it is essential to have both an accurate and a concise representation of the scene, making it convenient to study the problem in rate-distortion framework.

In one of the earliest attempts to fulfill the above constraint, a rate-distortion optimal 3D scene structure is estimated, by a joint optimization of the number of bits used to encode the extracted dense depth field, and the quality of the video frame rendered from the 3D structure, via a Markov random field formulation [1]. Despite the success of this approach, it is noted that the scene representation can benefit from further improvements.

2. DENSE 3-D SCENE REPRESENTATION

A dense 3D reconstruction can be described either by a point-based representation, as a depth-map that is defined on the same lattice with the reference frame, or by a volumetric representation, such as voxels, or by a parametric surface [10]. However, the former two ignore the scene geometry hence suffer from considerable redundancy [1]. This drawback is remedied in parametric surface based representations, by utilizing the fact that the constituent points of a scene are usually members of higher level geometric entities.

Planes offer distinct advantages as basic description elements in such representations, since man-made environments and many natural scenes can be well-approximated by planar patches. Besides, they can be parameterized with a small number of elements, therefore, efficiently represented. Finally, planes are algebraically easy to handle, providing considerable computational savings.

The considerable body of research on piecewise planar scene representations can be categorized into two. In the first class, a planar surface is fit onto an irregular 3D point cloud. The work of Schindler is a good example, in which the point cloud is partitioned into cells, and RANSAC is utilized to determine the dominant plane in each cell [3]. In [2], an equivalent procedure is described for finding the dominant homographies induced by scene planes, in a 2D correspondence set.

Since the benefits of piecewise planar scene representation are already known, there is a considerable body of research on this subject. The existing methods can be categorized into two main classes: Those defining the problem as fitting planes to an irregular 3-D point cloud, possibly obtained by a sparse 3-D reconstruction algorithm, and the others employing triangular meshes.

The second approach is characterized by the use of triangular meshes, specifically *Delaunay triangulation*, a procedure to connect the points in a set to form a triangular mesh with certain optimality properties [4]. There exist successful algorithms, such as [5], utilizing only the 3D location information of an irregular point cloud. Image-based-triangulation techniques also incorporate the intensity information. The basic algorithm employs edge swaps in a given 2D triangular mesh to minimize the prediction error of the intensity values of an image of a scene, acquired by a

[°] This work is funded by EC IST 6th Framework 3DTV NoE and partially funded by TÜBİTAK under Career Project 104E022

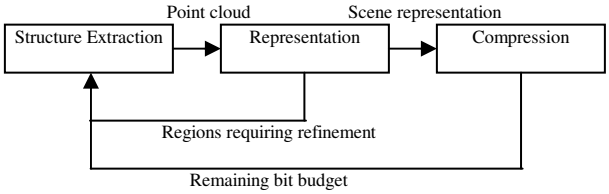

Figure 1: 3D reconstruction in rate-distortion framework.

known camera [6]. In [7], this method is improved by using simulated annealing to explore the solution space, enhanced by a rich arsenal of tools in addition to edge-swap. The method proposed in [8] uses a similar idea to compress a disparity map. However, it stands from the rest by its coarse-to-fine operation, adding new vertices to locations where the prediction error is largest.

Image-based triangulation methods construct the mesh on the 2-D projection of the 3-D point cloud, hence, suffer from erroneous connections. However, on the positive side, unlike irregularly-shaped planes generated by plane fitting process, a Delaunay triangulation can be represented solely by its vertices. Also, rendering of triangular meshes are supported by hardware. These advantages justify the choice of triangular meshes in this study.

In this paper, a simple, coarse-to-fine, 3-D piecewise planar reconstruction algorithm is proposed. The algorithm starts with a coarse mesh, two images of the scene and the corresponding cameras, and uses the intensity prediction error to drive the mesh refinement process.

3. RATE- DISTORTION EFFICIENT PIECEWISE PLANAR 3-D RECONSTRUCTION

3.1. Motivation

A rate-distortion efficient representation should provide, ideally, the least distortion possible for a given number of vertices. A sequential distortion minimization algorithm can proceed either in a fine-to-coarse, or a coarse-to-fine fashion. The former leads to a more complex cost function with local minima, and the computationally infeasible practice of extracting information only to discard later on. Coarse-to-fine approach avoids both, and in addition, is amenable to progressive coding and construction of scalable bit streams. Hence, in this study, a coarse-to-fine approach is adopted. Such an algorithm is identified by two rules, governing the *location* and the amount of *refinement*. The location is chosen to maximally reduce the distortion in the representation, while the amount is dictated by the available number of bits, by, for example, a compression block. These principles are depicted in Figure 1.

In order to develop an algorithm in a rate-distortion framework, exact definitions of rate and distortion are required. In this study, *rate* is defined as the number of vertices in the mesh. However, the choice of distortion metric is not straightforward. In the literature, PSNR of a predicted image is the most popular distortion metric, despite its oversensitivity to the geometric errors. Also, minimization of an image-based error introduces a projective distortion to the structure estimate in case of erroneous camera estimates. The alternative is geometry-based error metrics, assessing how well the point cloud is modelled by the current scene representation.

When accurate camera matrices are available, the minima of both of these metrics coincide. Otherwise, minimizing the image distortion transfers the error to the structure and vice versa. This observation explains the popularity of PSNR in novel view synthesis and image prediction problems [1].

3.2. Proposed Method

The proposed algorithm aims to achieve a rate-distortion efficient and accurate 3D scene geometry representation, with distortion measured as the prediction error of the intensity values of a target image from a reference image. The inputs to the algorithm are an initial mesh with typically 4-8 vertices to define a bounding polygon for the scene (required by the incremental Delaunay triangulation algorithm), a reference image, a target image for distortion measurement, and the corresponding projective camera matrices.

The first step is to identify a patch that requires refinement. To this aim, a prediction of the target image is constructed, by transferring the pixels in the reference image to the image plane of the camera corresponding to the target image. The pixels are transferred via the homographies induced by the planar patches in the mesh, representing the current 3D scene geometry estimate. The patch, whose corresponding region in the target has the largest intensity prediction error, is marked for refinement.

The next step is to determine the location of the 3D vertex to be added. To this aim, in both reference and target images, the regions corresponding to the projection of the patch to be refined are declared as region-of-interests (ROI). Then, in each ROI, a set of prominent features which will allow an accurate 3D position estimate are extracted by *Harris corner detector*. To find the matching features, guided matching is employed. This is a technique that makes use of the fact that *fundamental matrix* constrains the possible matches of a feature in an image to a line in the other image [11]. The fundamental matrix can be easily computed from the camera matrices, given as input. For each matching pair, there is a corresponding 3D vertex.

The vertex that is to be added to the mesh is chosen as the one, which has the least conformity to the current scene geometry estimated for the ROI. This can be measured by the distance of the 3D vertices to the plane. However, this metric has a geometric significance only when calibrated cameras are available. Use of a projective metric, such as

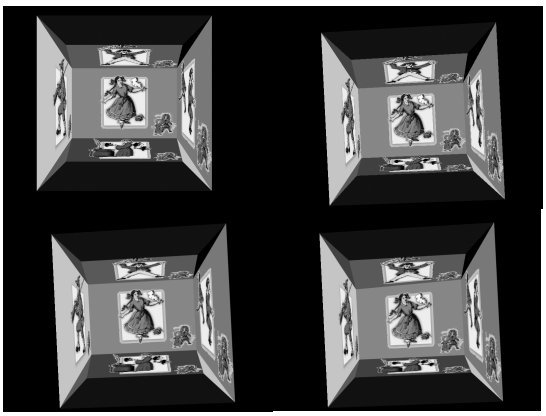

Figure 2: *"Cube". Upper-left:* Reference image. *Upper-Right:* Target image. *Lower-left*: Prediction, 4 points. *Lower-right:* Prediction, 11 points.

symmetric transfer error [11], removes this necessity. It is defined as

$$d = \left(\mathbf{x}_1 - \mathbf{H}^{-1}\mathbf{x}_1\right)^2 + \left(\mathbf{x}_2 - \mathbf{H}\mathbf{x}_2\right)^2 \quad (1)$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are the homogeneous coordinates of the matching pair and $\mathbf{H}$ is the homography induced by the planar patch, relating the ROIs.

The mesh is updated with the vertex which has the largest symmetric distance error. The procedure is repeated until either the intensity prediction error converges, or the available bit budget is completely used up. The flow of the algorithm is presented below.

`Algorithm: Piecewise-Planar Reconstruction`

Input: Initial (bounding) mesh, a reference image, a target image and the associated cameras

1. Until the prediction error converges or bit budget is depleted
2. Transfer the 2D points in the reference frame to the target frame, and compute the intensity prediction error in the regions corresponding to each planar patch.
3. Project the patch with the largest prediction error to the images to determine the ROIs. Extract new features and construct a correspondence set using guided matching in these ROIs.
4. Compute symmetric transfer error for each pair. Determine the pair with the largest transfer error, and add the corresponding 3D vertex to the mesh.
5. Go to Step 1.

Notice that, the algorithm minimizes the intensity prediction error, an image-based metric. However, new vertices are selected with respect to the symmetric transfer error, a geometric metric. Such an approach limits the distortion in the geometry, but obviously, the optimality of the chosen vertex is somewhat compromised.

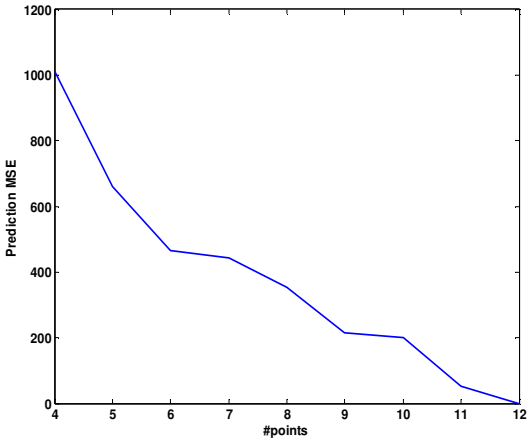

Figure 3: Rate-distortion curve for *"Cube"*

4. EXPERIMENTAL RESULTS

The algorithm is first tested on synthetic data, "*Cube"* (Figure 2), for which the ground-truth of the camera and geometry is available. Starting from a 4-point mesh, the algorithm successfully recovers all 12 points of the structure, while minimizing the prediction error during the process, as depicted in Figure 3. It should be noted that, depending on the constraints on the rate (i.e. number of points), 3-D recovery could be stopped at any point, i.e. the algorithm allows scalability.

Next, the algorithm is tested on *"Venus"* [12] (Figure 4), a data set for which only the uncalibrated cameras are known. The process, as depicted in Figure 6, starts with an 8-point reconstruction, and the prediction error converges at 40 points. The errors due to the automatic localization of the features and matching limit the residual error.

Finally, the algorithm is run on *"Cliff"*, a real sequence acquired from broadcast TV content (Figure 5), for which no information on cameras is available. The cameras, and the features are computed by the method described in [9]. The process starts with an 8-point mesh and the prediction error converges around 100 points (Figure 7). The increase in the residual error with respect to *"Venus"* can be attributed primarily to the errors in the camera matrices.

5. CONCLUSION

In this paper, a piecewise planar 3-D reconstruction algorithm is proposed. The algorithm starts with an initial coarse mesh, and seeks to achieve a favorable point in rate-distortion curve by refining the mesh by adding vertices to the worst planes, determined by the prediction error. The experiments indicate that the proposed algorithm can yield efficient representations, thus an important step towards rate-distortion optimal 3-D reconstruction. However, in

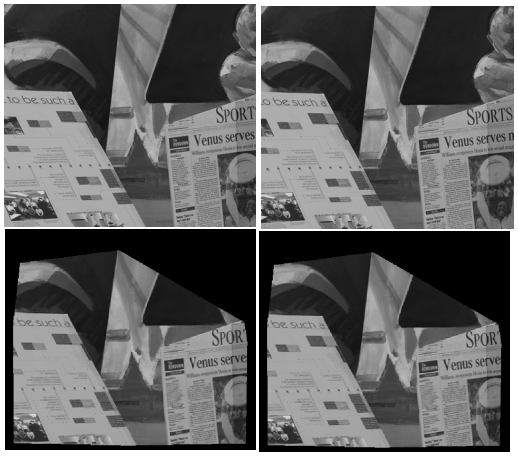

Figure 4: *"Venus". Upper-left:* Reference image. *Upper-Right:* Target image. *Lower-left*: Prediction, 8 points. *Lower-right:* Prediction, 40 points.

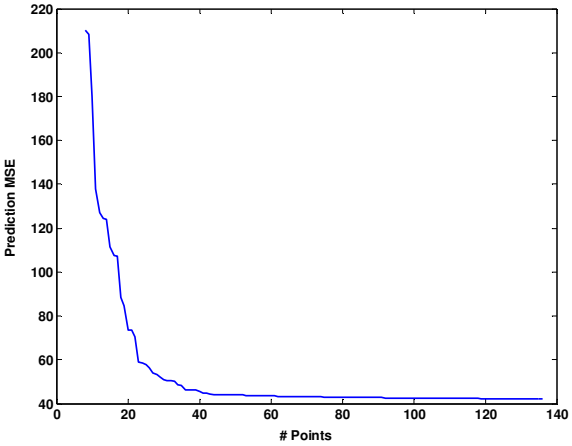

Figure 6: Rate-distortion curve for *"Venus"*

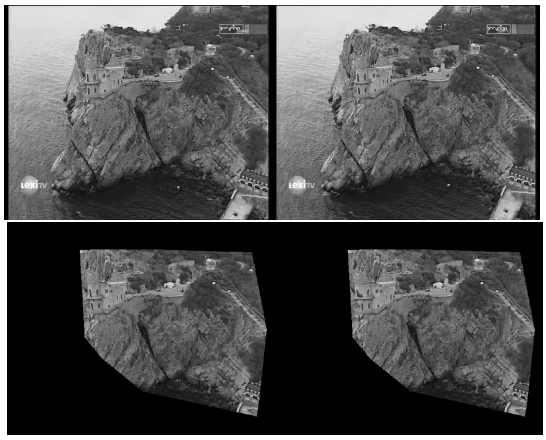

Figure 5: *"Cliff". Upper-left:* Reference image. *Upper-Right:* Target image. *Lower-left*: Prediction, 8 points. *Lower-right:* Prediction, 100 points.

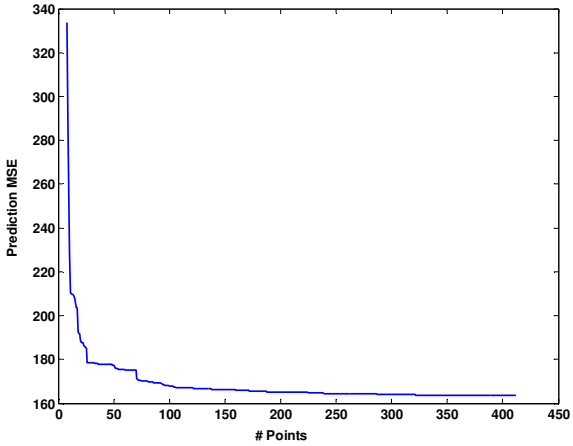

Figure 7: Rate-distortion curve for *"Cliff"*

applications in which camera and structure is not available and should be estimated from the sequence, the algorithm might face with some performance issues.

6. REFERENCES

[1] A Alatan, L.Onural, "Estimation of Depth Fields Suitable for Video Compression based on 3-D Structure and Motion of Objects," IEEE Trans. on Image Proc., no. 6, June 1998.

[2] A. Bartoli, P. Sturm, R. Horaud, "A Projective Framework for Structure and Motion Recovery from Two Views of a Piecewise Planar Scene", *Technical Report RR-4970,* 2000.

[3] K. Schindler, "Spatial Subdivision for Piecewise Planar Object Reconstruction*", Proc. of SPIE and IS&T Electronic Imaging- Videometrics VIII*, St. Clara, CA, 2003, 2003.

[4] R. Musin, "Properties of the Delaunay Triangulation", *Proc. of 13th Annual Symposium on Computational Geometry,* 1997.

[5] H. Hoppe, "Surface Reconstruction from Unorganized Points", *PhD. Thesis,* 1994.

[6] D. D. Morris, T. Kanade, "Image Consistent Surface Triangulation", *CVPR2000,* 2000.

[7] G. Vogiatzis, P. Torr, R. Cipolla, "Bayesian Stochastic Mesh Optimization for 3D Reconstruction", *BMVC2003,* 2003.

[8] J. H. Park, H. W. Park, "A Mesh Based Representation Method for View Interpolation and Stereo Image Compression", *IEEE Trans. on IP Vol.15, No.7,* 2006.

[9] E. Imre, S. Knorr, A. A. Alatan, T. Sikora, "Prioritized Sequential 3D Reconstructıon in Video Sequences of Dynamic Scenes", ICIP06, 2006.

[10] "3D Time Varying Scene Representation Technologies: A Survey*", 3DTV NoE Technical Report,* 2005.

[11] Hartley R., Zisserman A., "Multiple View Geometry in Computer Vision" ,Cambridge University Press, Cambirdge, 2003.

[12] http://cat.middlebury.edu/stereo/data.html