High Resolution Surface Reconstruction from Overlapping Multiple-Views*

Nader Salman†
INRIA Sophia Antipolis
France
Nader.Salman@sophia.inria.fr

Mariette Yvinec‡
INRIA Sophia Antipolis
France
Mariette.Yvinec@sophia.inria.fr

ABSTRACT

Extracting a computer model of a real scene from a sequence of views, is one of the most challenging and fundamental problems in computer vision. Stereo vision algorithms allow us to extract from the images a sparse 3D point cloud on the scene surfaces. However, computing an accurate mesh of the scene based on such poor quality data points (noise, sparsity) is very difficult. Here we describe a simple yet original approach that uses both the stereo vision extracted point cloud and the calibrated images. Our method is a three-stage process in which the first stage merges, filters and smoothes the input 3D points. The second stage builds for each calibrated image a triangular depth-map and fuses the set of depth-maps into a triangle soup that minimize violations of size and visibility constraints. Finally, a mesh is computed from the triangle soup using a reconstruction method that combines restricted Delaunay triangulation and Delaunay refinement.

Categories and Subject Descriptors: I.3.5 [Computational Geometry and Object Modeling]: Geometric algorithms,languages, and systems; I.3.8 [Applications]; I.4.5 [Reconstruction];

General Terms: Algorithms, Experimentation.

Keywords: Multiple-views, Surface reconstruction, Triangle soup, Restricted Delaunay triangulation, Delaunay refinement, CGAL , CIMG

1. INTRODUCTION

Three-dimensional reconstruction of real world scenes from sequences of overlapping images is a topic of much interest in Computer Vision, Computer Graphics and Computational Geometry. Its application areas range from preservation of historical heritage to e-commerce through computer animation, movie production and virtual reality walkthroughs.

In this video, we describe an approach to recover a computer model of a complex scene from a set of images taken from various view points. The models sought after are either simplicial surface meshes or 3D meshes of the objects in the scene.

Progress has been made in this field and working systems have been proposed. In the feature based stereo techniques, the first step usually involves matching, in pairs of input images, invariant feature detectors and descriptors. Then it is followed by camera auto-calibration, delivering camera pose information as well as a sparse 3D point reconstruction based on the image feature points. This may be followed by a further process in the images, which results in a per-pixel density (*i.e.* quasi-dense) 3D point reconstruction.

We assume that the acquisition process delivers sequences of located frames, *i.e.*, images with associated metadata describing for each image, the position of the camera as well as the shooting parameters. We use an automatic stereo vision algorithm to extract a set of tracks from the located frames, where each track is a 3D point associate to a list of frames where the point is visible. The set of images and tracks constitutes the input of the reconstruction algorithm.

Although the 3D points produced by the located frames processing step produce a quasi-dense sample, the point cloud is far from the quality of samples resulting from laser range scanners. The data points are entangled with redundancies, a large number of outliers and noise. Therefore, traditional 3D surface reconstruction techniques are not directly applicable to stereo data. Fortunately, the image information is still available which can be used to help surface reconstruction as seen in Lhuillier and Quan [10].

A first class of stereo vision reconstruction algorithms is based on the concept of visual hull [9]. These methods are mainly suited to single compact objects as their computational and memory cost becomes quickly prohibitive when the size of the scene increases. See Seitz et al. [15] for a review of some of the top-performing algorithms using visual hulls.

Volumetric methods, *i.e.*, methods based on a decomposition of the domain into elementary cells (grid, triangulations), include space carving [17, 5], level sets [13], and volumetric graph cuts [8].

Image based methods use image information (textures, camera position...) to help surface reconstruction. Bruzzone et al. [2] used a Delaunay triangulation constrained to the input feature line segments. The resulting two-dimensional triangulation in one of the images is then lifted into the 3D space, generating a triangular faced surface description. Morris and Kanade [11] described a scheme that searches for reasonable triangulations of a point set by considering the

---

*Work partially supported by the ANR *(Agence Nationale de la Recherche)* under the "Gyroviz" project (No ANR-07-AM-013)
http://www.gyroviz.org

†http://www-sop.inria.fr/members/Nader.Salman

‡http://www-sop.inria.fr/members/Mariette.Yvinec

SCG'09, June 8–10, 2009, Aarhus, Denmark.
ACM 978-1-60558-501-7/09/06.

textures in the images. Hilton [7] presented a reliable and computationally efficient recursive reconstruction algorithm which integrates the feature visibility independently for each view.

More recently, Strecha et al. [16] benchmarked multi-view stereo reconstruction methods which have proven to be adapted to reconstruct large scale building scenes. One of the principal similarities between these methods is the scene geometry representation by several depth maps.

We advocate for the combination of both volumetric and image based approaches. Furthermore, our method also combines stereo vision algortithms with methods used in surface reconstruction from point clouds. We argue that this particular synthesis of tools is the key to the efficiency of our method.

2. ALGORITHM DESCRIPTION

The algorithm takes as input the sequence of calibrated images provided with the corresponding set of tracks. We describe briefly the 3 steps of our algorithm.

Merging and filtering. Tracks whose 3D points are too close are merged in a single one with a list of camera updated to include the union of the cameras of the two merged tracks. Two criteria are then used to detect outliers in the set of tracks:
**** cone angle filtering -*** this criterion aims at eliminating the tracks which have been observed from only a few camera locations or from a set of camera directions forming small angles from the 3D point.
**** distance to neighbors -*** this criterion aims at eliminating tracks far away from densely populated regions of space. First we compute for each track p_i the average distance from p_i to its k-nearest neighbours, or $d_k(p_i)$ as follows:

$$d_k(p_i) = \frac{1}{k} \sum_{p_j \in \mathrm{kNN}(p_i)} \|p_i - p_j\|^2$$

Once the d_k of all tracks have been computed, we sort the tracks in order of increasing average squarred distance, and remove a small percentage of tracks with largest value.

Smoothing is then performed on the 3D points of the remaining tracks to eliminate noise.

Triangle soup. The algorithm then performs a contrast analysis [6] of the located frames in order to detect contour edges. These contours are either apparent contours in the images or the projection of sharp features of the scene. These contours are then used as constraints to build 2D Delaunay triangulations of the track projections in the image planes. Finally, a soup of 3D triangles (so-called "conformal soup") is obtained by lifting the 2D constrained triangulations in the 3D scene coordinates. By construction the triangles of this soup project onto homogeneous regions of the images in which they are seen. Two filtering steps using visibility constraint and size and shape criteria, are applied to the conformal triangle soup to remove erroneous triangles.

Reconstruction. In the reconstruction step, we compute a mesh from the triangle soup using a method that combine restricted Delaunay triangulation and Delaunay refinement [1, 14].

Our implementation is based on CGAL [3] for all the geometry processing and CIMG [4] for the image processing.

3. THE VIDEO

At first we show data sets from both feature based and dense-stereo algorithms. Then we show the different steps described in Section 2. Finally we show reconstruction results for various scenes for both sparse (Survey and Lion) and quasi-dense (Herz-Jesu-P25 [12]) point clouds.

The demonstration shown in the video uses OpenGL via MFC.

Acknowledgments. We wish to thank Laurent Saboret and Jean-Philippe Pons for their precious support.

4. REFERENCES

[1] J. Boissonnat and S. Oudot. Provably good sampling and meshing of surfaces. *Graphical Models*, 67(5):405–451, 2005.
[2] E. Bruzzone, M. Cazzanti, L. D. Floriani, and F. Mangili. Applying two-dimensional delaunay triangulation to stereo data interpolation. In *Proc. of the European Conference on Computer Vision*, pages 368–372, 1992.
[3] CGAL, Computational Geometry Algorithms Library. http://www.cgal.org.
[4] CIMG, C++ Template Image Processing Toolkit. http://cimg.sourceforge.net.
[5] O. D. Faugeras, E. Le Bras-Mehlman, and J. D. Boissonnat. Representing stereo data with the delaunay triangulation. *Artif. Intell.*, 44(1-2):41–87, 1990.
[6] W. Frei and C.-C. Chen. Fast boundary detection: A generalization and a new algorithm. *IEEE Trans. Comput.*, 26(10):988–998, 1977.
[7] A. Hilton. Scene modelling from sparse 3d data. *Image and Vision Computing*, 23(10):900 – 920, 2005.
[8] P. Labatut, J.-P. Pons, and R. Keriven. Efficient multi-view reconstruction of large-scale scenes using interest points, delaunay triangulation and graph cuts. *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8, Oct. 2007.
[9] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(2):150–162, 1994.
[10] M. Lhuillier and L. Quan. Surface reconstruction by integrating 3d and 2d data of multiple views. *Computer Vision, 2003. ICCV 2003. IEEE 9th International Conference on*, pages 1313, Washington, DC, USA, 2003.
[11] D. D. Morris and T. Kanade. Image-consistent surface triangulation. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:1332, 2000.
[12] MULTI-VIEW EVALUATION. http://cvlab.epfl.ch/data.
[13] J.-P. Pons, R. Keriven, and O. Faugeras. Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. *Int. J. Comput. Vision*, 72(2):179–193, 2007.
[14] L. Rineau and M. Yvinec. Meshing 3d domains bounded by piecewise smooth surfaces. In *Meshing Roundtable conference proceedings*, pages 443–460, 2007.
[15] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. *Computer Vision and Pattern Recognition, 2006. CVPR 2006. IEEE Conference on*, pages 519–528, Washington, DC, USA, 2006.
[16] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, June 2008.
[17] A. Treuille, A. Hertzmann, and S. Seitz. Example-based stereo with general BRDFs. In *Proc. of the European Conference on Computer Vision*, pages 457–469, 2004.