Physically Valid Triangulation of Sparsely Matched Images Using Texture Information: Application to View-Synthesis*

J.S. Perrier, G. Agam, and P. Cohen

Perception and Robotics Laboratory
Department of Electrical and Computer Engineering
Ecole Polytechnique, Montreal H3C 3A7, Canada

Abstract

Image-based view synthesis is an emerging research topic, with numerous applications. The ability to synthesize new views from existing images, enables the generation of enhanced and synthetic environments. Approaches for view synthesis relay upon dense or sparse matching of the views. In both cases, some parts of the images are inevitably unmatched. Triangulation of sparsely matched points present a possible solution for the handling of those unmatched regions. However, such triangulation should respect the underlying geometry of the scene.

In this paper, a novel approach is proposed for physically valid triangulation of sparsely matched points. The proposed approach is based on the maximization of a physical validity criterion which is supported by textured regions in the images. The produced triangulation is such that each triangle corresponds approximately to a planar surface in the scene. Given an arbitrary initial triangulation, the proposed approach refines it by flipping the edges of triangles. Furthermore, since missing matched points may preclude the correct triangulation of the scene, an additional stage handles the addition of matched points inside low score triangles. Inherent to our approach, the support region which is used for the evaluation of the correctness of the added matches is normally much larger than the one used in a local match evaluation. The paper contains examples of image-based view synthesis of real and virtual scenes, produced by the proposed approach.

1 Introduction

In recent years, image-based view synthesis has become the focus of intensive research for applications that range from the efficient photo-realistic rendering of complex virtual or real scenes to the efficient compression of video sequences. The different approaches to image-based view synthesis could be classified into methods that include an explicit 3D reconstruction or some other complete representation, and methods that do so only implicitly. The various methods differ in their calibration needs, the required number and closeness of the input views, the density and accuracy of required matched points between the views, and the way in which the information is re-projected to create new views. When the input views are strongly calibrated and densely matched, a complete 3D model of the scene can be estimated and used to generate arbitrary new views. Arbitrary views may also be produced from a voxel [11] or plenoptic [8] representations, which could be generated based upon a dense set of calibrated and densely matched input views. With three densely matched input views, the trilinear tensor can be retrieved to generate new views by either reconstructing a 3D model or re-projecting the input views directly [2]. With two densely matched input views, the fundamental matrix can be computed and used to generate new views by re-projecting the pixels of the input views onto a new image plane [6]. Finally, without calibration, by using only a set of matched points between two images, view interpolation and view morphing techniques [4, 10, 12] can be used to generate new images. A thorough review and classification of image-based view synthesis techniques may be found in [1].

While many methods for image-based view synthesis are based upon the establishment of dense matchings between corresponding views, the sole reliance on densely matched points is sensitive to local match errors and cannot produce results for uniform (non-textured) regions in which dense matching cannot be reliably obtained. Consequently, such methods do not necessarily guarantee improved results with respect to methods which are based on a sparse, cleverly selected set of matched points. Moreover, sparse matchings may be obtained more efficiently and with greater reliability by using robust techniques. Scenes, particularly man-made

*This work is supported by the IRIS-3 Network of Centers of Excellence (Project SMART NR-COH).