Chapter 8

Monocular–Cues Based 3–D Reconstruction: A Comparative Review

Sudheer Tumu
State University of New York-Albany, USA

Viswanath Avasarala
GE Global Research, USA

Sai Tejaswi Jonnalagadda
Hetero Med Solutions, India

Prasad Wadekar
Mahindra Satyam, USA

ABSTRACT

3-D reconstruction from images has traditionally focused on using multiple images. However, in recent years, some interesting breakthroughs have been made in constructing depth maps of images using monocular cues. This chapter summarizes the recent work in this evolving research domain. The chapter also presents results from an initial exploratory study based on fusing the 3-D reconstructions generated by two seminal monocular-cue based reconstruction algorithms. With new testing data, the fusion approach improved the 3-D estimation accuracy significantly as compared to the original approaches. The authors use the improved estimation accuracy produced by the fusion algorithm as a motivating evidence for future work: the use of non-parametric Bayesian regression for 3-D reconstruction.

INTRODUCTION

Traditionally, computational vision algorithms for constructing 3D structure from images have focused on using multiple images of a scene. Based on triangulation, these techniques including ‘Structure from Motion (SFM)’ or ‘Stereovision’ calculate the depth maps of images. These approaches are not applicable to situations where only single images of a particular scene are available. On the other hand, humans are adept at analyzing the 3D structure of even a single
image. For this purpose, human brain relies on a number of monocular cues including properties related to image texture, known object sizes/shapes, occlusion, vanishing points etc. Many of these monocular cues depend not only on the local properties of image but also it’s global properties and are therefore difficult to model. Because of the difficulty in understanding/modeling these monocular cues, automating the identification of the 3D structure from a single image is a difficult problem. However, in recent years, researchers have made significant progress in developing solutions for this problem. Particularly, the work done by Saxena et al. (Saxena, A., Sun, M., Ng, A. Y., 2009; Saxena, A., Chung, S.H., Ng, A. Y., 2007, & Saxena, A., Cheung, S.H., Ng, A. Y., 2005) and Hoiem et al. (Hoiem, D., Efros, A., & Herbert, M., 2005 & Hoiem, D., Efros, A., & Herbert, M., 2006) has significantly improved the state of the art in calculating 3D structure from monocular cues. Both these approaches rely on sophisticated machine learning algorithms to learn the feature patterns useful for estimating 3D structure of images. In this chapter, we summarize the two above seminal approaches to 3D construction from monocular cues and discuss their comparative performance. We also present results of our fusion algorithm that combines the two results to produce better depth estimates. Section II provides a brief literature review of both monocular and multi-view based 3-D reconstruction approaches. Section III describes the Saxena et al. (2009)’s Make3D algorithm and Hoiem et al. (2006)’s geometric image labeling work (referred to as HEH in the rest of this chapter) in greater detail. We describe the design of our fusion algorithm and results from an experimental study in the results section. We are currently investigating non-parametric Bayesian approaches for monocular depth estimation. This approach is introduced in the future work session.

BACKGROUND

The 3-D reconstruction problem can be formulated in multiple ways (Zhang, R., Tsai, P.-S., Cryer, J., & Shah, M., 1999). A common formulation requires the calculation of depth value at each coordinate. The depths can be expressed as the distance of the surface point to the camera or from any other standard surface like the x-y plane. Other ways to express 3-D structure is by using surface normals, surface gradients or surface slant and tilt angles. In addition to these quantitative approaches, 3-D reconstruction algorithms can simply generate geometric/qualitative labels that allow the organization of the image into a 3-D hierarchy.

In this section, we provide a high level review of previous work on 3-D reconstruction both using binocular or multiple views and monocular views.

3-D Reconstruction from Multiple Images

Many techniques for constructing 3-D geometry based on two or more images of a scene have been successfully developed. Stereovision algorithms are the most common 3-D reconstruction algorithms.

As a first step, cameras are calibrated and image rectification is performed so that image planes become coplanar. Correspondence algorithms are used to match stereo pairs between images (Faugeras, O., & Luong, Q.T., 2001; Hartley, R. I., & Zisserman, A., 2000). The correspondence algorithms are usually pixel-based (refer to Scharstein, D., & Szeliski, R., 2002 for a good survey). Frequency domain-based correspondence algorithms, particularly using Gabor Filters or FFTs (Udo Ahlvers, U.Z, 2005) exhibit sub-pixel accuracy and are also used. Depth values or the inversely related disparity values of pixels can be calculated using epipolar constraints (other approaches are described in Scharstein, D., et al., 2002). The disparity calculation uses a minimization procedure that minimizes the aggregate...
14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:

www.igi-global.com/chapter/monocular-cues-based-reconstruction/64390?camid=4v1

This title is available in InfoSci-Intelligent Technologies, InfoSci-Books, Science, Engineering, and Information Technology, InfoSci-Computer Science and Information Technology, InfoSci-Select. Recommend this product to your librarian:

www.igi-global.com/e-resources/library-recommendation/?id=16

Related Content

Recognition of Chemical Entities using Pattern Matching and Functional Group Classification
www.igi-global.com/article/recognition-of-chemical-entities-using-pattern-matching-and-functional-group-classification/171439?camid=4v1a

Supporting Empathetic Boundary Spanning in Participatory Workshops with Scenarios and Personas
www.igi-global.com/article/supporting-empathetic-boundary-spanning-participatory/74368?camid=4v1a

Color Image Segmentation of Endoscopic and Microscopic Images for Abnormality Detection in Esophagus
www.igi-global.com/chapter/color-image-segmentation-endoscopic-microscopic/62690?camid=4v1a

Fuzzy Logic-Based Intelligent Control System for Active Ankle Foot Orthosis
www.igi-global.com/chapter/fuzzy-logic-based-intelligent-control-system-for-active-ankle-foot-orthosis/178438?camid=4v1a