

Chapter 13

Modular Stereo Vision: Model and Implementation

Ng Oon-Ee

Monash University Sunway Campus, Malaysia

Velappa Ganapathy

University of Malaya, Malaysia

S.G. Ponnambalam

Monash University Sunway Campus, Malaysia

ABSTRACT

The two-frame stereo vision algorithm is typically conceived of and implemented as a single process. The standard practice is to categorize individual algorithms according to the ‘type’ of process used. Evaluation is done based on the quality of the depth map produced. In this chapter, we demonstrate that the stereo vision process is actually composed of a number of inter-linked processes. Stereo vision is shown to be modular in nature; algorithms implementing it typically implement distinct stages of the entire process. The modularity of stereo vision implies that the specific methods used in different algorithms can be combined to produce new algorithms. We present a model describing stereo vision in a modular manner. We also provide examples of the stereo vision process being implemented in a modular manner, with practical example code. The purpose of this chapter is to present this model and implementation for the use of researchers in the field of computational stereo vision.

INTRODUCTION

“Stereo vision” is a specific method for 3-dimensional imaging. Other methods include multi-view vision, optical flow, and photometric methods relying on known light sources or shadows. Stereo vision is quite robust in application, not relying on

any fixed external constraints (besides the scene being lighted). Stereo vision setups do not require controlled illumination which may disturb the scene being observed. They are relatively cheap in cost and power consumption relative to multi-view methods and laser rangefinder imaging.

The disadvantages of stereo vision are high complexity and ambiguity in the stereo vision process itself (this is expanded on in the next

DOI: 10.4018/978-1-61350-326-3.ch013

section). ‘Pure’ two-frame stereo vision is much more difficult than multi-view methods which can rely on many different observations; it is therefore more vulnerable to noise and lack of scene texture. Much research has been done into solving the stereo vision problem; some of that work is referenced in this chapter.

In this chapter, we examine the stereo vision process as a system composed of distinct processes. The stereo vision process is typically viewed as a kind of black box which describes a single process applied on two input images to produce a depth map. This chapter describes the processes which make up the typical stereo vision algorithm. Based on these processes, a model for modular stereo vision is presented.

This model is not meant merely as an academic description, but as a practical guide to implementing stereo vision algorithms. Algorithms developed based on this model will have component parts (called modules) which are interchangeable with other implementations of the same process. In this way, a newly developed approach for one process in the model can potentially be used in any number of existing algorithms. To this end, this chapter also presents a practical implementation of the model in C++.

To summarize, this chapter presents a model describing the stereo vision process as a series of inter-linked processes. The practical advantages of this model over the standard ‘black box’ view of stereo vision are discussed. An implementation of this model is included to demonstrate those advantages.

COMPUTATIONAL STEREO VISION

This section provides a brief introduction to computational stereo vision. Readers who are familiar with stereo vision and its implementation on computers should feel free to skip to the next section on Modular Stereo Vision.

The first sub-section describes the challenges inherent in stereo vision, and gives a general summary of how these challenges are tackled by existing algorithms. The second sub-section discusses how the typical stereo vision algorithm is considered to be a black box when compared against a competing algorithm and the background behind this understanding. The final sub-section briefly discusses the quantitative measurement of stereo vision result quality.

The Stereo Vision Challenge

Stereo vision is the process of converting two views of a scene to its depth map. For computational stereo vision, typically two cameras are used in the epipolar configuration with parallel visual axes (Figure 1). This is analogous to the physical arrangement of our eyes (Ganapathy and Ng, 2008, Figure 1). In general, the term ‘stereo vision’ should only be used to describe the two-view case, since multi-view vision is a very different field of research. Also, some hybrid depth-perception methods rely on a combination of visual data with other non-visual cues; these are not sufficiently related to stereo vision to warrant inclusion in this discussion.

The depth of a scene element is measured from the baseline between the camera apertures. The disparity of a scene element is the difference between the position of that element in each view. For example, the disparity d of Object 1 in Figure 1 is $x_{l_1} - x_{r_1}$. Given a baseline b between the two cameras and a common focal length f , the relationship between depth z and disparity is:

$$z = \frac{b \cdot f}{d} \quad (0.1)$$

Due to this inverse relationship, if we know the disparity of an object, we can calculate its depth and therefore its position relative to the cameras. The real challenge in stereo vision is matching

21 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/modular-stereo-vision/60269

Related Content

An Efficient Generalized Error Concealment in Video Codec

Ansari Vaqar Ahmed and Uday Pandit Khot (2020). *International Journal of Computer Vision and Image Processing* (pp. 1-28).

www.irma-international.org/article/an-efficient-generalized-error-concealment-in-video-codec/264218

Inexpensive, Simple and Quick Photorealistic 3DCG Modeling

Ippei Torii, Yousuke Okada, Manabu Onogi and Naohiro Ishii (2013). *Image Processing: Concepts, Methodologies, Tools, and Applications* (pp. 550-561).

www.irma-international.org/chapter/inexpensive-simple-quick-photorealistic-3dgc/77562

Assessment of Kidney Function Using Dynamic Contrast Enhanced MRI Techniques

Melih S. Aslan, Hossam Abd El Munim, Aly A. Farag and Mohamed Abou El-Ghar (2010). *Biomedical Image Analysis and Machine Learning Technologies: Applications and Techniques* (pp. 214-233).

www.irma-international.org/chapter/assessment-kidney-function-using-dynamic/39562

A State-of-the-Art Survey on Face Recognition Methods

Prashant Modi and Sanjay Patel (2022). *International Journal of Computer Vision and Image Processing* (pp. 1-19).

www.irma-international.org/article/a-state-of-the-art-survey-on-face-recognition-methods/283961

A Robust Skeletonization Method for Topological Complex Shapes

Donatella Giuliani (2017). *International Journal of Computer Vision and Image Processing* (pp. 1-18).

www.irma-international.org/article/a-robust-skeletonization-method-for-topological-complex-shapes/177198