Chapter 21
Stereo Vision Depth Estimation Methods for Robotic Applications

Lazaros Nalpantidis
Royal Institute of Technology (KTH), Sweden

Antonios Gasteratos
Democritus University of Thrace, Greece

ABSTRACT

Vision is undoubtedly the most important sense for humans. Apart from many other low and higher level perception tasks, stereo vision has been proven to provide remarkable results when it comes to depth estimation. As a result, stereo vision is a rather popular and prosperous subject among the computer and machine vision research community. Moreover, the evolution of robotics and the demand for vision-based autonomous behaviors has posed new challenges that need to be tackled. Autonomous operation of robots in real working environments, given limited resources requires effective stereo vision algorithms. This chapter presents suitable depth estimation methods based on stereo vision and discusses potential robotic applications.

INTRODUCTION

Stereo vision is a reliable tool in order to exploit depth data from a scene, apart the pictorial one. The accuracy of the results depends on the choice of the stereo camera system and the stereo correspondence algorithm. Stereo correspondence is a flourishing field, attracting the attention of many researchers (Forsyth & Ponce, 2002; Hartley & Zisserman, 2004). A stereo correspondence algorithm matches pixels of one image (reference) to pixels of the other image (target) and returns

DOI: 10.4018/978-1-61350-326-3.ch021

the corresponding vertical displacement as the reference pixel's disparity, which is proportional to its depth. Thus, stereo vision is able to retrieve the third dimension of a scenery and, therefore, its importance is obvious in issues such as traversability estimation, robot navigation, simultaneous localization and mapping (SLAM), as well as in many other aspects of production, security, defense, exploration and entertainment.

Stereo correspondence algorithms can be grouped into those producing sparse and those giving dense output. Feature based methods stem from human vision studies and are based on matching segments or edges between two images, thus resulting in a sparse output. This disadvantage is counterbalanced by the accuracy and the speed of calculations. However, robotic applications demand more and more dense output. This is the reason why most of the relevant literature is focused on stereo correspondence algorithms that produce dense output. In order to categorize and evaluate them a context has been proposed (Scharstein & Szeliski, 2002). According to this, dense matching algorithms are classified in local and global ones. Local methods (area-based) trade accuracy for speed. They are also referred to as window-based methods because disparity computation at a given point depends only on intensity values within a finite support window. Global methods (energy-based) on the other hand are more time consuming but very accurate. Their goal is to minimize a global cost function, which combines data and smoothness terms, taking into account the whole image. Of course, there are many other methods that are not strictly included in either of these two broad classes. A detailed taxonomy and presentation of dense stereo correspondence algorithms can be found in (Scharstein & Szeliski, 2002). Additionally, the recent advances in the field as well as the aspect of hardware implementable stereo algorithms are covered in (Nalpantidis, Sirakoulis, & Gasteratos, 2008b).

ISSUES OF ROBOTICS-ORIENTED STEREO VISION

While a heavily investigated problem, stereo correspondence is far from being solved. Furthermore, the recent advances in robotics and related technologies have placed more challenges and stricter requirements to the issue. However, common problems related to outdoor exploration, such as possible decalibration of the stereo system and tolerance to non-perfect lighting conditions, have been barely addressed. Robotic applications demand stereo correspondence algorithms to be able to cope with not ideally captured images of the working environments of the robots (see Figure 1) and at the same time to be able to provide accurate results operating in real-time frame rates. Some of the open issues of robotics-oriented stereo vision methods are the handling of non-ideal lighting conditions, the requirement for simple calculation schemes, the use of multi-view stereo systems, the handling of miscalibrated image sensors, and the introduction of new biologically inspired methods to robotic vision.

Non-Ideal Lighting Conditions

The correctness of stereo correspondence algorithms' depth estimations is based on the assumption that the same feature in the two stereo images should have ideally the same intensity. However, this assumption is often not valid. Even in the case that the gains of the two cameras are perfectly tuned, so as to result in the same intensity for the same features in both images, the fact that the two cameras shoot from a different pose, might result in different intensities for the same point, due to shading reasons. In general, stereo image pairs captured in real life environments often suffer from differentiations in illumination, as those shown in Figure 2. Moreover, in real environments, which is the case for robotic applications, the illumination is far from being ideal (Klancar, Kristan, & Karba, 2004; Hogue, German, & Jenkin, 2007).