Chapter 21
Stereo Vision–Based Object Matching, Detection, and Tracking:
A Review

Mohamed Saifuddin
Sunway University, Malaysia

Seng Kah Phooi
Sunway University, Malaysia

Lee Seng Yeong
Sunway University, Malaysia

Ang Li-Minn
Edith Cowan University, Australia

ABSTRACT

Computer vision has become very important in recent years. It is no longer restricted to a single camera that is only capable of capturing a single image at any given time. In its place, stereo vision systems have been introduced that not only make use of dual cameras to capture multiple images at once, but they also simulate the exact same nature of the human eye vision. Stereo vision has turned out to be an important research component in the subdivision of computer vision and image processing that deals with the extraction of information from images for the purpose of video surveillance systems, mimicking the human vision for the visually impaired, for robotics, to control unmanned vehicles, for security purposes, virtual reality and 3 Dimensional (3D) televisions, etc. In this chapter, a comprehensive review of all recent algorithms such as stereo matching, object detection, tracking techniques for stereo vision are presented.

INTRODUCTION

Computer vision is a subject, which attempts to recreate the human vision by fabricating models which seem to have related properties to visual observation. It also tried to develop models which seem to have related characteristics to graphical perceptions. Stereo vision is an essential component of computer vision. It is the extraction of 3D information from digital images by examining the relative positions of the objects captured by the dual cameras. With the help of stereo vision, it is possible to reconstruct,

either partially or fully, a 3D scene from two or more images that have been captured under marginally dissimilar angles. There are two main categories of computer vision: plane and stereo vision. The most notable difference between them is the depth information i.e. distance of the objects from the cameras. This is not possible to be detected by a single camera as it requires the use of dual cameras. Each lens captures its own view and then two independent images are sent to the system for processing. The system compares the images while shifting the two images together over top of each other to find the parts that match. The shifted amount is called disparity. The disparity at which objects in the image best match is used by the system to calculate their distance. When both the images are processed by the system, they are combined into a single image by matching up the similarities and then adding in the minor differences. These minor differences between the captured images combine to give out a relatively bigger difference in the resultant image. This combined image is greater than the aggregate of its parts. It is a 3D stereo image.

Innovative work in the field of Computer Vision began in the early 60's when Robert from MIT had successfully finalized a 3D scene analysis project(Weiss, 1999). In his project, in order to achieve the 3D scene analysis, 2D image processing had been employed. This particular project had long been considered to be the origin of the stereo vision technique. Nowadays, a comprehensive stereo vision system could be created by concealing the initial steps from capturing images to the final step of recreating the visual surface of the objects. The concept of 'computation stereo' was first proposed by Barnard and Fisher(Arnaud, 2004) who explained that it covered the topics of image matching, depth information, image acquisition and also feature extraction. This has led to the conclusion that a stereo vision system with dual cameras can be utilized at the same time to capture the left & right images effectively so as to obtain the required depth information to be used in a range of applications such as a video surveillance system, mimicking the human vision for the visually impaired, for robotics, to control unmanned vehicles, for security purposes, virtual reality and 3D TV etc.

In a stereo vision process, the primitives extracted from the images that are being matched such as segments, pixels, regions etc. is found to be the most significant phase. There are two extensive types of matching methods (Banks, Bennamoun, Kubik, & Corke, 1997) .The first one makes use of pixel neighborhood correlation method, which generates a dense disparity map, whereas the other method makes use of matching based on characteristics, in this case, generating a disparity map that is comparatively sparse. In any case, the stereo matching based on edge points utilizing linear images is given more importance.

Figure 1 illustrates the general processes of stereo vision based object detection and tracking. In comparison to single camera systems, a stereo vision system is able to overcome the main problems associated with the reconstruction and correspondence issues (TKanade & Okutomi, 1993). Initially, the images are acquired from the left and right cameras which then undergo three stages of processing: stereo-matching, object detection and object tracking. Each camera seems to capture its own perception of the image and the two separate images are then transferred to the acquisition state for further processing. During the stereo matching processing, the two input images are compared with each other in order to find the parts that match. These images are then merged into a single image by pairing the similarities and combining the marginal differences together. This difference between the two images is known as disparity and it is used to identify the distance between the two images. There are two main types of stereo matching algorithm called dense and sparse disparity algorithms (Nalpantidis, Georgios, & Antonios, 2008). The generated disparity map is filtered with the depth-based background modeling process in order to remove shadows. With the depth information, the objects are detected on the plan-view map. The stereo frames and the depth information are used to detect the 3D points cloud. Then, the

Related Content

A Conceptual and Pragmatic Review of Regression Analysis for Predictive Analytics
Sema A. Kalaian, Rafa M. Kasimand Nabeel R. Kasim (2017). *Decision Management: Concepts, Methodologies, Tools, and Applications (pp. 1761-1776).*
www.irma-international.org/chapter/a-conceptual-and-pragmatic-review-of-regression-analysis-for-predictive-analytics/176831

Estimating the Global Demand of Photovoltaic System
Yi-Fen Chen, Bi-Chu Chen, Chia-Wen Tsai, Wen-Yu Chenand Lee-Wei Wei (2012). *International Journal of Strategic Decision Sciences (pp. 120-128).*
www.irma-international.org/article/estimating-global-demand-photovoltaic-system/63659

Information Supporting of Decision Making for Energy Management in District Heating
Vira Shendryk, Victor Nenia, Olga Aleksenkoand Yuliia Parfenenko (2017). *Optimum Decision Making in Asset Management (pp. 310-333).*
www.irma-international.org/chapter/information-supporting-of-decision-making-for-energy-management-in-district-heating/164058

Marketing Modeling and Validation
 (2012). *Systems Thinking and Process Dynamics for Marketing Systems: Technologies and Applications for Decision Management (pp. 198-223).*
www.irma-international.org/chapter/marketing-modeling-validation/65307

A Regression Dependent Iterative Algorithm for Optimizing Top-K Selection in Simulation Query Language
Susan Farley, Alexander Brodskyand Chun-Hung Chen (2012). *International Journal of Decision Support System Technology (pp. 12-24).*
www.irma-international.org/article/regression-dependent-iterative-algorithm-optimizing/70653