Chapter 15
Stereo Vision in a Network of Co-operative Cameras

Sanjeev Kumar, Christian Micheloni, and Gian Luca Foresti

Abstract In this chapter, a smart framework is presented for the object localization on a given ground-plane test map using heterogeneous stereo vision. In particular, the image pairs are captured by using static and Pan-tilt-zoom (PTZ) cameras, which are heterogeneous in terms of imaging parameters thus having different focal lengths, image resolutions, intensities, etc. These two cameras are selected in a cooperative manner from a network of static and PTZ cameras and used as a stereo system to localize an object even in the case when it is partially occluded. The various sequences of images captured by these cameras are made homogeneous based on their focal ratio and then by performing zero padding. The pairs of matching points are obtained using scale invariant features (SIFT) matching from stereo images. The rectification transformations are calculated by solving a constrained nonlinear optimization problem. The 3-D position of the object is estimated based on a modified concept of stereo matching in rectified pairs of images. Localization is made using the 3-D position obtained from stereo. Experiments are performed to evaluate the performance of the proposed framework using real sequences of images. The proposed method is useful in stereoscopic as well as in video surveillance applications.

15.1 Introduction

Modern video surveillance systems are more intelligent and smarter than conventional passive video monitoring systems in terms of design and applicability. A number of steps are involved in video surveillance to monitor a wide and complex area such as target detection, localization, and tracking. Localization of moving objects on a given ground-plane map is an important and crucial task in video surveillance systems. In general, it is not practical to cover a wide area with different cameras in addition to the possibility of missing information when there are significant gaps
in coverage. As a solution, researchers have developed some techniques for video surveillance using dynamic cameras or hybridization of static–dynamic cameras. Generally, these kinds of techniques are based on a 2-D homography between the camera plane and a given ground-plane test map. However, these techniques are not able to perform accurate localization when the object is partially occluded. Smart and adaptive technologies are needed to overcome the above drawback and perform the task of localization as accurately as possible.

Stereo vision has the advantage that it is able to estimate an accurate and detailed 3-D representation of the position of an object with respect to a given co-ordinate system using its two or more perspective images. Generally, a pair of identical cameras is used in classical stereo systems. These cameras are placed along a horizontal line near each other. The 3-D positions of the objects are estimated by solving a geometrical inverse problem, in other words, by using the inverse solution of collinearity equations. These collinearity or perspective projection equations describe the relation between the 3-D position of the objects and their 2-D co-ordinates in the image plane of the camera.

Nowadays, a large number of cameras are used to monitor a wide area. Hence, these cameras can be used as stereo system to achieve effective surveillance. Usually, the distance between these cameras cannot always be kept small as in classical stereo systems. If the two cameras are placed far apart from each other then classical stereo algorithms fail to provide the required 3-D position of the objects. The other main disadvantages are the lack of co-operation between these sensors and heterogeneity in the imaging parameters. These difficulties can be tackled by introducing a smart algorithm. On the other hand, there are many advantages such as more degrees of freedom in terms of field of view (Field of view (FOV)) and angle of view (Angle of view (AOV)) by using a stereo unit with heterogeneous cameras (usually a pair of static and PTZ cameras). More accurate results can be obtained by applying these heterogeneous cameras based on stereo systems for localization of objects.

### 15.1.1 Related Research

In most of the research reported in the literature, a pair of identical cameras is used to build a stereo system [89, 186]. Many real-life problems related to computer vision are solved using these identical camera-based stereo systems only. Generally, these cameras cover fixed fields of view (FOVs). In surveillance applications, one or both identical cameras are replaced by dynamic or PTZ cameras to design a more flexible stereo system [557]. In this way, one can get more degrees of freedom in terms of FOV and AOV. To monitor a large and complex area, a combination of static and PTZ cameras can be used to obtain better results [305].

Recently, research has been carried out by using the hybridization of fixed and PTZ cameras or dual PTZ cameras in stereo vision [305, 557]. Some intelligent approaches related to hybrid-sensors surveillance systems have been proposed in detail [5, 225, 443]. The system proposed in [5] has been successfully applied in