2014 IEEE International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)
October 18-19, 2014, Wuhan, Hubei, China

A Shape-based Stereo Matching Algorithm for Binocular Vision

Xinjian Fan, Xuelin Wang and Yongfei Xiao
Shandong Provincial Key Laboratory of Robot and Manufacturing Automation Technology
Institute of Automation, Shandong Academy of Science
Jinan, Shandong, P.R.China
e-mail: fanxj@keylab.net, wxl0036@yahoo.com, fei9818@163.com

***Abstract*—Binocular stereo vision is an important branch of the research area in computer vision. Stereo matching is the most important process in binocular vision. In this paper, a new stereo matching scheme using shape-based matching (SBM) is presented to improve the depth reconstruction method of binocular stereo vision systems. The method works in two steps. First, an operator registers the pattern including the key features of an object to be measured. Then during the operation stage, the stereo camera snaps stereo images and finds the patterns in right and left images separately by means of the SBM. The 3D positions of the object are calculated by using the corresponding points of the stereo images and the projection matrices of the stereo camera. Since we apply robust image processing algorithms, such as the SBM, the proposed method becomes more reliable than the conventional stereo vision systems.**

Keywords—binocular vision; stereo matching; shape-based matching; object detection

I. INTRODUCTION

Binocular stereo vision is an important branch of the research area in computer vision. It directly simulates the manner of human eyes observing a scene from two different viewpoints. A binocular stereo vision system commonly uses two cameras that are separated by a short distance and capture the scene simultaneously. Based on the principle of triangulation, the 3D information of a point can be recovered by its vision disparity from two images.

The most important process in binocular vision is stereo matching which identifies the corresponding points between two images. Up to now a great deal of computer vision research has been addressed to this problem and there are many approaches on it. The existing techniques for stereo matching are roughly grouped into two categories. One is intensity-based methods and the other is feature-based methods [1]. Intensity-based methods use dense low-level features and intensity. Accordingly, these methods are sensitive to noise and small intensity differences caused by illumination variations or projective distortion [2]. The feature-based methods rely on finding sparse features such as Harris corners [3], SIFT features [4], zero-crossing points [5], edges [6], line segments, etc. Compared to intensity-based methods, these methods are generally robust to changes in illumination, noise, occlusion and minor changes in viewpoint. However, feature-based methods also have limitations due to feature instability.

In this paper, to overcome the traditional problems of stereo matching, we propose a new stereo matching method in which correspondence points are determined using shape-based matching (SBM) [7]. SBM compares the shape features between the current input image and the pre-registered pattern. Contrary to the traditional pattern matching method commonly known as "normalized correlation", SBM makes use of a set of shape features instead of the intensity, so it robustly finds the target object despite variations in size, angle, and illumination. SBM has proved to be one of the most powerful pattern matching methods, especially in industrial applications. It seems that SBM serves as the key technology at the heart of some commercial machine-vision packages, such as HALCON's shape-based matching [8].

The method works as follows. In the preparatory stage, an operator registers the pattern including the key features of an object and trains the registered pattern to get a model. When in the operation stage, the stereo camera snaps stereo images and finds the patterns in right and left images separately by means of the SBM. Then 3D positions of the object are calculated by using the corresponding points of the stereo images and the projection matrices of the stereo camera. Since we apply robust image processing algorithms, such as the SBM, the proposed vision system becomes more reliable than the conventional stereo vision systems.

II. SYSTEM OVERVIEW

Fig. 1 describes the architecture of the proposed algorithm. It can be divided into a preparatory stage and an operation stage. In the preparatory stage, the pattern of each object has to be registered and the binocular sensor calibrated. In the operation stage, the object to be measured is first detected using the SBM method, which compares the shape features between the current input image and the pre-registered pattern in the database. This object detection is performed in both images of the stereo camera. According to its 2D positions, 3D coordinates of the object is then calculated through the determination of a correspondence point in the stereo image and triangulation. The main technical functions of the proposed algorithm are camera calibration, object detection, and 3D coordinate reconstruction. They will be explained in detail in the following sections.

This work was supported by Shandong Provincial S&T Development Project under Grant No. 2013GGX10313 and Shandong Provincial Natural Science Foundation under Grant No. ZR2013FM026.

978-1-4799-5353-0/14/$31.00 ©2014 IEEE

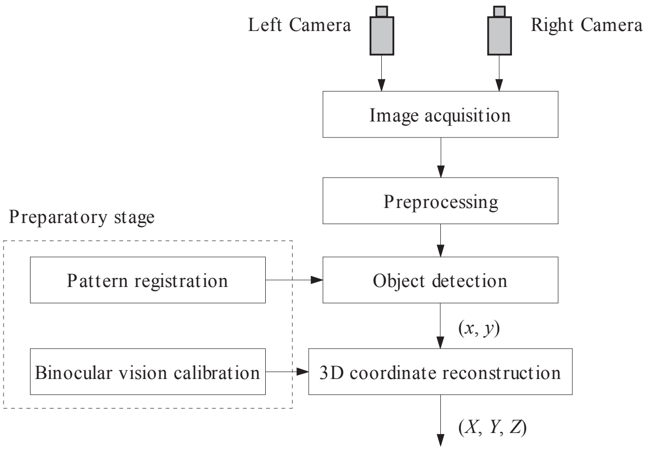

Fig. 1. The system workflow

(a) Before rectification (b) After rectification

Fig. 2. Input image rectification

III. Camera Calibration, Image Acquisition and Preprocessing

A. Binocular Vision Calibration

The calibration of a binocular vision system mainly includes two parts:

1) One is single camera calibration, which is to obtain the intrinsic parameters, extrinsic parameters and distortion coefficients. It has been widely discussed [9][10].

2) The other part is structure parameters calibration, which aims to determine the structure parameters that describe the spatial relationship between the two cameras.

Thus the whole process can roughly be divided into two steps. At the first stage, the single camera calibration is conducted by means of the well-known Zhang's method [9]. After that, the structure parameters which represent the relative position and rotation of the two cameras are determined by optimization methods.

B. Image Acquisition

Acquiring 2D images is the first step to obtain the 3D information of an object in binocular stereo vision. Two cameras capture the same scene at the same time. The output of this step is stereo images. In this paper, the 640×480 size images are used for scanning.

C. Preprocessing

1) Low-pass filtering

In order to rectify the images it is important to smooth them using low-pass filtering. Otherwise the rectified images will exhibit aliasing effects.

2) Rectification

Rectification is the process of correcting input images for the distortions of the lenses. Lenses often cause distortions in raw images. Further, rectified images will be corrected so that the rows of images digitized from horizontally displaced cameras are aligned, and similarly that the columns of images obtained from vertically displaced cameras are aligned. Alignment is computed using rotation and translation parameters from two calibrated cameras. As a result, the images of the two cameras are re-projected so that they reside exactly in the same plane. Without this feature, searching along the rows and columns will not produce the correct results. Fig. 2 demonstrates the results of image rectification.

IV. Object Detection Using Shape-based Matching

A. Shape-based Matching

Pattern matching can be subdivided into two categories, depending on the type of input patterns: correlation-based matching and feature-based matching [11]. Correlation-based methods work directly on images which are considered as arrays of intensity values or real valued functions. Feature-based methods work on geometric data such as finite point sets or polygons. This geometric data may be obtained from images using feature extraction techniques.

Shape-based matching (SBM) is a known feature-based method for 2D object recognition [7][12]. As its name suggests, objects are represented and recognized by their shape. Here, the shape is extracted by selecting all those points whose contrast exceeds a certain threshold; typically, the points correspond to the contours of the object. Contrary to the traditional correlation-based methods, SBM makes use of a set of geometric features instead of the intensity, so it robustly finds the target object despite illumination variations, noise, clutter and partial occlusion. In this paper, we use a commercial implementation of SBM available in the HALCON library [8]. Halcon's superior SBM can localize objects with sub-pixel accuracy in realtime, even if they are rotated and scaled.

B. The Pattern Registration

To find the correspondence points by using the SBM, the pattern of an object to be measured should be pre-registered. An operator registers the pattern including the key features of the object and trains the registered pattern to get a model. Fig. 3 shows the pattern of a circular flange. It consists of one main feature that involves the inner hole of the object, and four sub-features, which are small circles distributed uniformly on the workpiece.

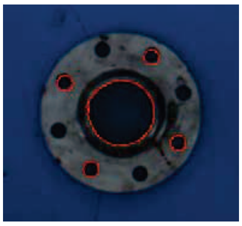

Fig. 3. Registered pattern of a workpiece

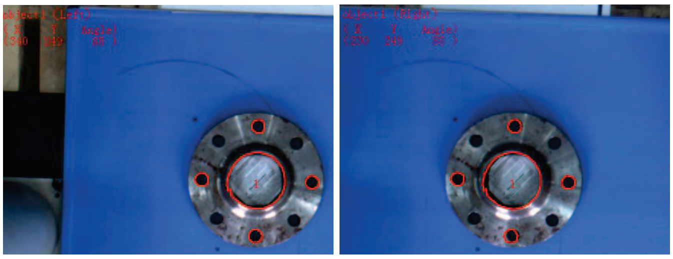

Fig. 4. Object detection results in *left* image and *right* image

C. Object Detection

Given an input image, the aim of object detection is to give the location, orientation and size of an object, if any. It is performed by comparing an input window and the object model on every position of an image pyramid of the input image. Fig. 4 is the result of the object detection for the stereo images. The image coordinates of the detected patterns in the object detection process are used as the correspondence points of the stereo image.

V. 3D Coordinate Reconstruction

If the projection matrices of two cameras are given, and if the pair of corresponding points in the stereo images through the object detection process is given, we can determine the coordinate of the scene point which corresponds to the points in the two image plane. Fig. 5 explains the geometry of binocular stereo vision systems.

The relationship between a point (X, Y, Z) in the scene and its corresponding points (u_l, v_l) and (u_r, v_r) in the two camera coordinates can be written as follows:

$$\begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} \cong \mathbf{M}_l \bullet \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11}^l & m_{12}^l & m_{13}^l & m_{14}^l \\ m_{21}^l & m_{22}^l & m_{23}^l & m_{24}^l \\ m_{31}^l & m_{32}^l & m_{33}^l & m_{34}^l \end{bmatrix} \bullet \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} \cong \mathbf{M}_r \bullet \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11}^r & m_{12}^r & m_{13}^r & m_{14}^r \\ m_{21}^r & m_{22}^r & m_{23}^r & m_{24}^r \\ m_{31}^r & m_{32}^r & m_{33}^r & m_{34}^r \end{bmatrix} \bullet \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{2}$$

where $\mathbf{M}_l$ and $\mathbf{M}_r$ denote the projection matrix of the left camera and the right camera respectively. Equations (1) and (2) are expressed in terms of homogeneous coordinates:

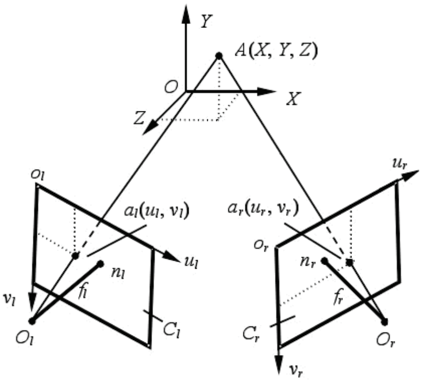

Fig. 5. Geometry of binocular cameras for 3D reconstruction

$$\mathbf{A} \bullet \mathbf{X} = \mathbf{B} \tag{3}$$

where

$$\mathbf{A} = \begin{bmatrix} (u_l m_{31}^l - m_{11}^l) & (u_l m_{32}^l - m_{12}^l) & (u_l m_{33}^l - m_{13}^l) \\ (v_l m_{31}^l - m_{21}^l) & (v_l m_{32}^l - m_{22}^l) & (v_l m_{33}^l - m_{23}^l) \\ (u_r m_{31}^r - m_{11}^r) & (u_r m_{32}^r - m_{12}^r) & (u_r m_{33}^r - m_{13}^r) \\ (v_r m_{31}^r - m_{21}^r) & (v_r m_{32}^r - m_{22}^r) & (v_r m_{33}^r - m_{23}^r) \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix},$$

$$\mathbf{B} = \begin{bmatrix} m_{14}^l - u_l m_{34}^l \\ m_{24}^l - v_l m_{34}^l \\ m_{14}^r - u_r m_{34}^r \\ m_{24}^r - v_l m_{34}^r \end{bmatrix}$$

Therefore we can reconstruct 3D points $\mathbf{X}$ using pseudoinverse as shown in (4).

$$\mathbf{X} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}}\mathbf{B} \tag{4}$$

In the case that the optical axes of two cameras are parallel (so-called *parallel-optical-axes binocular cameras*), $\mathbf{X}$ can be simply computed by disparity d:

$$\mathbf{X} = \frac{b}{d}\mathbf{C} \tag{5}$$

where $d = u_l - u_r$, b is the baseline distance, $\mathbf{C} = \begin{bmatrix} u_l \\ v \\ f \end{bmatrix}$, $v_l = v_r = v$, $f_l = f_r = f$.

VI. Experimental Results

Our experiment setup is shown in Fig. 6. A Bumblebee2 camera, made by Point Grey Research (PGR), is used as the vision hardware. The stereo camera has two Sony ICX204 CCD cameras and the optical axes of two cameras are parallel. The base line of the stereo camera is 12 cm. The camera connects via an IEEE 1394 link to a PC. A commercial circular flange is used as a reference object for measuring. The position of the object is denoted by the 3D coordinates of its center. The stand-off distance between the camera and the object was about

Fig. 6. Experimental setup

0.5m to 0.8m (minimum and maximum distances). The vision processing system was developed under Microsoft visual studio and it runs on an IPC with an Intel Core2Duo P8700 CPU (2.54Ghz) and 3GB RAM.

To evaluate the performance of the proposed method, we conducted experiments that measured 3D positions of the workpiece in different spatial places. In each place, we ran our algorithm for 5 times. The average results are compared with those gotten by using stereo matching functions provided by the PGR company, which are based on the Sum of Absolute Differences (SAD) correlation method. The results by the PGR functions can be considered as accurate ones, for the positioning error is less than 0.5mm when the stand-off distance is within 800mm [13]. Table 1 gives some of the experimental data. From Table 1, it can be seen that the SBM method achieves similar matching accuracy compared to the SAD method. The maximum positioning difference by the two methods is 0.60% in Z axis, 2.62% in X axis and 3.25% in Y axis. The difference in X-Y direction becomes large when X-coordinate and Y-coordinate approach to the optical axis (that is, X-coordinate and Y-coordinate become small). We analyzed that for the same $\Delta X = X_{sbm}-X_{sad}$, $DX = \Delta X/X_{sad}$ becomes large when X_{sad} becomes small. The reported processing time is 292ms for the SBM method and 410ms for the SAD method on the average. Apparently, our method works faster than the SAD method. SAD may fail to establish correspondence of the pixels in the stereo images due to the presence of noise or illumination variations. Since SBM is a very robust and powerful image processing algorithm, we think that the proposed method is more reliable than most of the conventional stereo matching methods.

VII. Conclusion

A new stereo matching algorithm for binocular vision is proposed in the paper. In order to improve the 3D positioning performance, the SBM method is used to find the correspondence points in the stereo images. In contrast with the conventional binocular vision system, it could easily calculate the 3D position of the target object by just registering the patterns with respect to the features in the object. Therefore it could be employed in various general purpose applications without the limitations of the shape and the material of the object. Through experiments on various commercial industrial workpieces, we validated that the proposed method accurately measured the 3D position of a randomly placed workpiece on the pallet. Since the proposed method combines the process of object-of-interest detection and stereo matching in a natural way, it is particularly suitable for industrial applications such as bin-picking or visual servoing. In the near future, we will do further research on 1) further improving the 3D positioning accuracy by combining other mature stereo matching methods; 2) a structured-light technique for dealing with a textureless object; 3) pose estimation using SBM.

References

[1] R.M. Haralick and L.G. Shapiro, Image matching, Computer and Robot Vision, Part 2, Addison-Wesley, New York, pp. 289-371, 1992.

[2] C. L. Zitnick and T. Kanade, "A cooperative algorithm for stereo matching and occlusion detection," IEEE Trans. Pattern Anal. Mach Intell., vol. 22, pp. 675-684, 2000.

[3] C.Harris and M. Stephens, "A combined corner and edge detector," Proc. 4th Alvey Vision Conf., pp. 147-151, 1988.

[4] D. G. Lowe, "Object recognition from local scale-invariant features,"

TABLE I. Experimental Results of 3D Positioning (unit: mm)

SBM Matching			**SAD Matching**			**Positioning difference[a]**		
X	*Y*	*Z*	*X*	*Y*	*Z*	*DX*	*DY*	*DZ*
-244.3	42.7	719.3	-246.8	42.4	723.6	1.00	0.68	0.60
56.7	44.9	739.0	55.9	43.5	737.0	1.35	3.25	0.27
56.4	234.7	734.9	55.7	234.0	733.8	1.24	0.30	0.15
-244.3	229.9	718.1	-245.5	229.5	719.6	0.48	0.16	0.22
-245.5	-98.7	723.8	-246.9	-100.1	723.8	0.54	1.44	0.00
57.5	-101.2	737.4	56.0	-102.1	737.7	2.62	0.81	0.03
74.1	41.0	552.6	73.8	40.6	554.5	0.42	1.13	0.33
-227.2	40.7	544.1	-228.3	39.8	544.5	0.49	2.25	0.07
-227.0	-115.3	543.1	-227.9	-115.6	543.0	0.38	0.26	0.02
74.5	-117.3	554.8	73.5	-117.9	553.9	1.29	0.52	0.17
73.8	152.4	555.1	72.6	151.5	553.5	1.72	0.59	0.29
-227.5	148.8	542.9	-228.1	148.1	543.5	0.23	0.42	0.11

a. Computed by 100*|SBM-SAD|/SAD.

Proc. 7th IEEE Int'l Conf. on Computer Vision (ICCV'99), pp. 1150-1157, Corfu, Greece, 1999.

[5] W. Eric and L. Grimson, "Computational experiments with a feature based stereo algorithm," IEEE Trans. Pattern Anal. Mach Intell., vol. 7, no. 1, pp. 17-34, 1985.

[6] G. Medioni and R. Nevatia, "Segment-based stereo matching," Computer Vision, Graphics, and Image Processing, vol. 31, pp. 2-18, 1985.

[7] C. Steger, "Similarity measures for occlusion, clutter, and illumination invariant object recognition," Proc. 23rd DAGM-Symposium on Pattern Recognition, pp. 148-154. Springer-Verlag, 2001.

[8] MVTec, HALCON Version 10.0 – Solution Guide II-B: Matching, 2010.

[9] Z. Y. Zhang, "A flexible new technique for camera calibration," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 11, pp. 1330-1334, 2000.

[10] J. Heikkila, "Geometric camera calibration using circular control points," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 10, pp. 1066-1077, 2000.

[11] Michiel Hagedoorn, Pattern Matching Using Similarity Measures, PhD thesis, Universiteit Utrecht, 2000.

[12] C. Steger, "Occlusion, clutter, and illumination invariant object recognition," Int'l Archives of Photogrammetry and Remote Sensing, vol. XXXIV, Part 3A, pp. 345–350, 2002.

[13] Point Grey Research, Point Grey Research Stereo Accuracy, 2010.