A study of stereo microscope measurements based on interpolated feature matching

Yigang Wang[a,b], Gangyi Jiang[a,*], Mei Yu[a], Shengli Fan[a,b] and Jing Deng[a]
[a]*Faculty of Information Science and Engineering, Ningbo University, No. 818, Fenghua Road, Ningbo, 315201, China*
[b]*Department of Information, Ningbo Institute of Technology, Zhejiang University, No. 1, Qianhu South Road, Ningbo, 315010, China*
[c]*School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast Ashby Building, Stranmillis Road, Belfast BT9 5AH, UK*

Abstract. Recently, digital stereo microscopes have been becoming increasingly popular in fields such as biomedicine and biotechnology. As such, being able to precisely match left/right feature pairs is critical, which is an important step in stereo matching. We hereby propose a dynamic interpolation method that combines two dimensional calibration and depth information to achieve stereo matching. In this experiment, we obtained the original scale-invariant feature transform feature point pairs and selected a pair of matching units. Based on the selected pair of matching elements and interpolated feature points shown in the left view of stereo microscope image, the dynamically interpolated feature points in the right view were developed to perform stereo measurements. The experimental results revealed that the proposed method can realize stereo microscopic measurements; results with high reliability can be achieved with a large quantity of matching elements.

Keywords: Dynamically interpolated feature points, stereo matching, stereo matching, digital stereo microscope

1. Introduction

Digital stereo microscopic measurements match the coordinates of the World Coordinate System (WCS) with corresponding matching points that are derived from matching left and right views of stereo image. These useful tools have become increasingly popular in the field of biomedicine [1-3]. Several stereo matching methods have been applied in different cases. In cases that high precision is required, (such as stereo measurements and image mosaics), and high-precision matching methods that reach sub-pixel levels are applied. In other cases (such as disparity generation and stereo coding), block matching and region matching approaches are adopted to achieve matching efficiency through cost function. Considering the above requirements, stereo matching is divided into two categories: those based on features and those based on regions. For stereo matching approaches based on features, the common features are Scale Invariant Feature Transform (SIFT) features and geometric features. In

---

order to improve the efficiency of the stereo matching of high-speed automobile navigation, Sharma et al. combined SIFT features with maps before being stored in the database [4, 5]. Joglekar, et al. reported relevant parameters of neural networks based on rigorous SIFT matching, and achieved denser stereo matching with the neural networks based on Bayesian decision [6]. Yang, et al. studied the applications of SIFT in wide baseline stereo matching, and improved the matching accuracy by eliminating mismatched points through baseline constraint based on SIFT matching [7].

With the geometric features, Lee, et al. applied stereo matching to high-resolution satellite images of the South Pole through geometric constraint matching [8]. Zeng, et al. applied the high-resolution camera to building measurements in which the geometric position of the roof and the color matching of the stereo images were combined with geometric matching to measure parameters such as building height [9]. In traditional Chinese medical science, locations should be determined during robot massage in which accurate geometric and color features are difficult to obtain due to the particularity of human skin. However, Wang, et al. adopted a geometric center of gravity to assist in locating, which gave better results [10]. Conventional stereo matching can be achieved by cost function. Higher-order cost function is applied for stereo matching on the discontinuous edge in which better effects are achieved [11]. Antunes, et al. achieved effective stereo measurements by using a new cost function by establishing a virtual plane in which data relevance is guaranteed by symmetric clues [12]. Kim, et al. presented a Mahalanobis distance method based on cross correlation in which robust stereo matching is achieved in different light exposure levels [13]. Robust stereo matching was also achieved in different light exposure levels and colors by Nguyen, et al [14].

The quantity and accuracy of left/right view points of stereo image have direct influences on the ultimate results. In this paper, for stereo microscope measurements on IC chips, the conventional approach results in errors because the quantity of matched feature point pairs of left/right view is little by its shape rules. Therefore, a new method is proposed in this paper, in which dynamic feature point pairs are generated to measure depths of IC chips' grooves.

2. Stereo microscope measurement method based on interpolated feature matching

2.1. Overall algorithm process

Figure 1 shows the flowchart of the proposed algorithm. We first combine the object and calibration board to obtain the stereo microscopic images. Then, stereo calibration is performed based on local images of calibration board by using a depth compensation approach [15]. Pre-calibration is achieved by collecting pictures of calibration boards at different depths and default CMOS parameters. Then, pre-calibrated parameters are corrected based on depth information to achieve camera calibration. The SIFT features in the image region including the object [16] is extracted and stereo matching is accomplished using the minimum Euclidean distance. After that, the matching elements are selected from the matching points while the mismatching pairs are eliminated. The matching point pairs are dynamically generated. Finally, the depth of IC chips' groove is measured via stereo reconstruction based on all matched point pairs.

2.2. Camera calibration

Stereo calibration is a key element for the ultimate measurements. All calibration methods are based on pinhole imaging shown in Figure 2 and Eq. (1). In these methods, the intrinsic and extrinsic
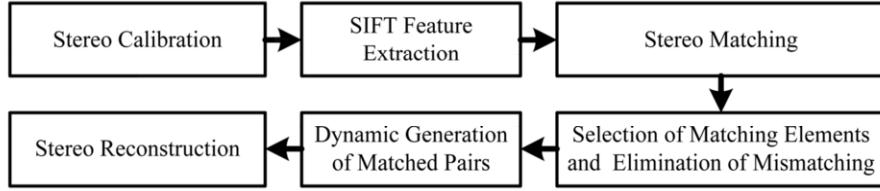
Fig. 1. A flowchart of stereo measurements system based on interpolated matched features.
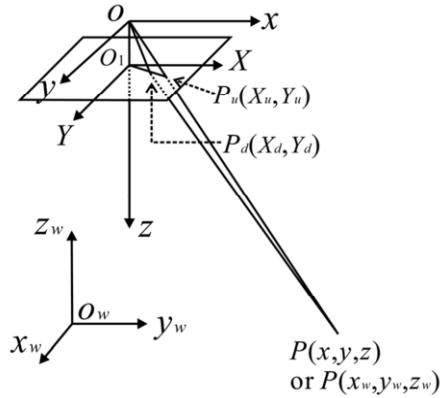

Fig. 2. Pinhole imaging of the camera.

parameters of the camera are obtained using the calibration points. (X_i, Y_i) is the coordinate of pixels and (x_w, y_w, z_w) is the coordinate of any point in WCS. $\mathbf{M}_1$ and $\mathbf{M}_2$ represent the intrinsic and extrinsic parameters of the camera, respectively. z is the Z-axis coordinate of the camera coordinate system of (x_w, y_w, z_w) (in mm). (X_u, Y_u) is the ideal projection coordinate of point P on the camera imaging plane according to pinhole imaging theory. (X_d, Y_d) is the actual projection coordinate on the camera imaging plane when considering factors such as lens distortion. This study does not address the distortion issue.

$$z\begin{pmatrix} X_i & Y_i & 1 \end{pmatrix}^T = \mathbf{M}_1 \mathbf{M}_2 \begin{pmatrix} x_w & y_w & z_w & 1 \end{pmatrix}^T \tag{1}$$

We use this stereo microscopic calibration method with compensation for depth information because the depth of the microscope inhibited calibration through conventional means. Let z_w and z'_w denote calculated depth and actual depth respectively, then $z'_w = kz_w$. The ultimate results can be corrected using this equation, as there is a linear relationship between the calculated depth according to the predicted parameters of the camera and actual depth.

2.3. Matching elements selection and mismatching elimination

For pinpoint positioning, SIFT features are selected as matching elements. The 4x4 block with 8 different directions is selected as the description. Therefore, a total of 128 components are used to describe the feature. Firstly, matched points of the left and right views can be obtained via minimum
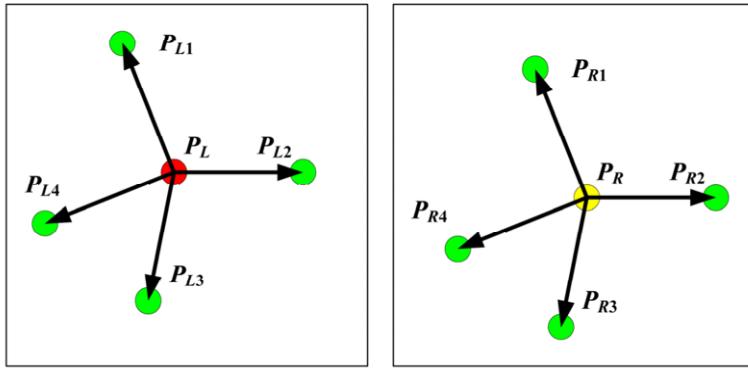
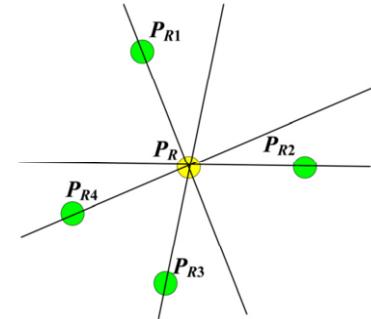Fig. 3. Illustration of point pairs being dynamically interpolated.

Fig. 4. The practical situation of dynamic interpolation.

Euclidean distance. The descriptors for the left and right images are $Desp_L$ and $Desp_R$, respectively.

$$Desp_R = \min_j (\sum_{n=1}^{128} (Desp_L[n] - Desp_{jR}[n])^2) \qquad (2)$$

Therefore, after obtaining all matched point pairs, RANSAC (random sample consensus algorithm) is utilized to eliminate the apparent mismatching point pairs. Afterwards, K point pairs are randomly selected among the remaining point pairs as the matched elements for subsequent interpolation.

2.4. Interpolation process of matched feature point pairs

For the IC chip images involved in this paper the SIFT feature point pairs are limited as all images showed large-size smooth regions except for the apparent geometric features. In order to achieve stereo microscopic measurements with better robustness, feature points are dynamically interpolated. The principle is shown in Figure 3. In the above figure, the green points are the obtained matched elements while $(P_{L1}, P_{R1}), (P_{L2}, P_{R2}), (P_{L3}, P_{R3})$ and (P_{L4}, P_{R4}) are the matched point pairs.

The selection of matched point pairs is based on the principles that three points on the same line should be avoided and the area of convex hull generated by matching units should be maximized. The red points P_L are interpolated dynamic feature points. The yellow points P_R are feature points dynamically matched with P_L, which are predicted according to P_L and matched elements. There are a total of K pairs of matched elements, while Figure 3 shows 4 pairs for illustration. According to Figure 3, P_R satisfies the following requirements: $\overrightarrow{P_L P_{L1}} = \overrightarrow{P_R P_{R1}}$, $\overrightarrow{P_L P_{L2}} = \overrightarrow{P_R P_{R2}}$, ..., $\overrightarrow{P_L P_{LK}} = \overrightarrow{P_R P_{RK}}$.

P_R can be solved by equalizing any vector mentioned above. Multiple sets of data are combined to guarantee the accuracy and validity of the ultimate outcome. Considering the complexity of the algorithm, vector directions are used to solve the equations, as shown in Figure 4. Since the K lines are less likely to pass by the same point, the least square method is used to obtain P_R.

The coordinates of P_{Li} are (x_{Li}, y_{Li}), and the coordinates of P_{Ri} are (x_{Ri}, y_{Ri}). Additionally, $i = 1, 2, \cdots, K$. The coordinates of P_L and P_R are (x_L, y_L) and (x_R, y_R), respectively. Since the directions are identical, the following equation can be obtained by:

$$(x_{Li} - x_L)/(y_{Li} - y_L) = (x_{Ri} - x_R)/(y_{Ri} - y_R) \tag{3}$$

We arrange the information to get:

$$(y_{Li} - y_L \quad -(x_{Li} - x_L))(x_R \quad y_R)^T = (y_{Li} - y_L)x_{Ri} - (x_{Li} - x_L)*y_{Ri} \tag{4}$$

(x_R, y_R) can be gotten by applying linear regression to the following Eq. (5) through the LSM.

$$\begin{cases} (y_{L1} - y_L \quad -(x_{L1} - x_L))(x_R \quad y_R)^T = (y_{L1} - y_L)x_{R1} - (x_{L1} - x_L)y_{R1} \\ (y_{L2} - y_L \quad -(x_{L2} - x_L))(x_R \quad y_R)^T = (y_{L2} - y_L)x_{R2} - (x_{L2} - x_L)y_{R2} \\ \vdots \\ (y_{LK} - y_L \quad -(x_{LK} - x_L))(x_R \quad y_R)^T = (y_{LK} - y_L)x_{RK} - (x_{LK} - x_L)y_{RK} \end{cases} \tag{5}$$

2.5. Calculation of world coordinate system

After stereo matching, the WCS can be calculated by combining the matched points with the camera parameters obtained via prior calibration. (x_L, y_L) and (x_R, y_R) are the coordinates of p_L and p_R, which are the midpoints in Figure 5. P is the spatial point in both cases. (x_w, y_w, z_w) is the coordinate of spatial point P in the WCS. Its value can be obtained based on the following Eqs. (6) and (7). Herein, $\mathbf{M}_L$ and $\mathbf{M}_R$ are the obtained intrinsic and extrinsic matrix after calibration.

$$z_{cL}\begin{pmatrix} x_L \\ y_L \\ 1 \end{pmatrix} = \mathbf{M}_L\begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix} = \begin{pmatrix} m_{11}^L & m_{12}^L & m_{13}^L & m_{14}^L \\ m_{21}^L & m_{22}^L & m_{23}^L & m_{24}^L \\ m_{31}^L & m_{32}^L & m_{33}^L & m_{34}^L \end{pmatrix}\begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix} \tag{6}$$

$$z_{cR}\begin{pmatrix} x_R \\ y_R \\ 1 \end{pmatrix} = \mathbf{M}_R\begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix} = \begin{pmatrix} m_{11}^R & m_{12}^R & m_{13}^R & m_{14}^R \\ m_{21}^R & m_{22}^R & m_{23}^R & m_{24}^R \\ m_{31}^R & m_{32}^R & m_{33}^R & m_{34}^R \end{pmatrix}\begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix} \tag{7}$$

3. Results and discussions

3.1. Equipment and methods

The stereo microscope (ZOOM460N) used was purchased from Nanjing Wave-optics Pte. Ltd. As shown in Figure 6(a), the two separate cameras (A) were to capture left and right images, (B) was the main body of the microscope, and (C) was the sample platform. (D1), (D2), and (D3) were the motors controlling the movement of the sample platform in three dimensions. The highly precise lead screw used in this equipment could control the movement of the platform to levels of 2 μm.

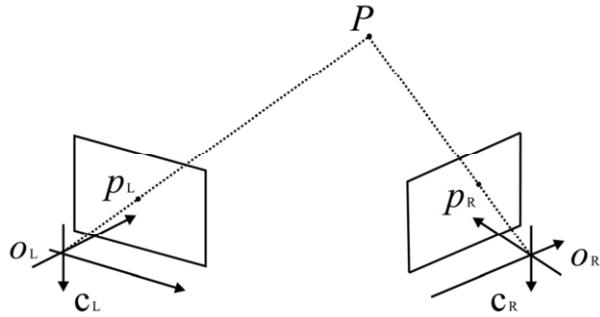The stereo effects could be achieved by the two individual cameras capturing images of the object

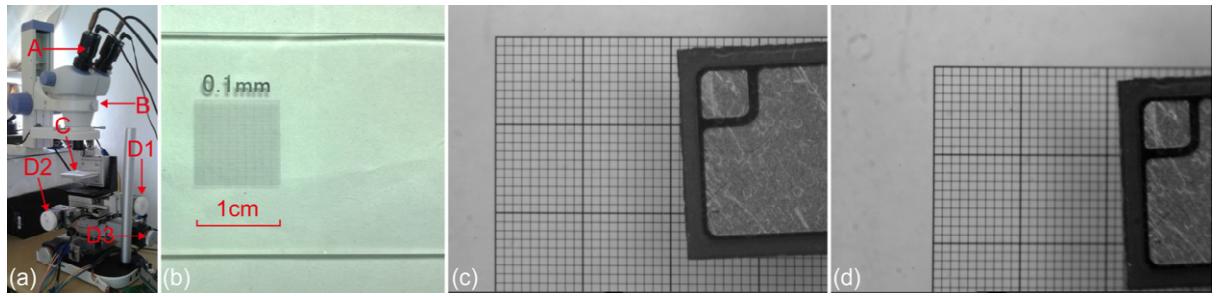Fig. 5. Illustration of stereo reconstruction.

Fig. 6. (a) is stereo microscope, (b) is scaling board, (c) is left image captured from left camera and (d) is right image.

separately. The cameras (Watec 902B) were placed at the location of the eyepiece lens; images (704*576 per frame) were obtained via USB sampling. Scaling was achieved using 10 mm * 10 mm scaling board (100 x 100, each square unit is 0.1 mm * 0.1 mm), as shown in Figure 6(b).

In this experiment, the depth of grooves in power IC devices was measured. The magnification of the microscope used is 0.7. The chips were placed on the scaling board; left and right images were obtained separately. Then, the platform was moved in the Z-direction by 1 step (2 μm) and the images were obtained again. The images used for analysis were taken at the 90th (- 0.06 mm), 120th (origin) and 150th (+ 0.06 mm) step. Figures 6(c) and 6(d) shows the images obtained at the 120th step.

3.2. Selection of scaled image pair and scaling results

This process was done according to previous reports [15]. With the images obtained in the previous step as the benchmark, the results of camera scaling are as follows. Where the linear depth rectification coefficient $k = 2.1605$. Here, the focal lengths of the left and right cameras exhibited a significant variation. This can be attributed to the difference in the locations of these two cameras and the amplification effect by the object lens.

$f_L = 104.141665$, $f_R = 649.191868$, $\kappa_L = -0.00006356341$, $\kappa_R = 0.0001940109$.

$\mathbf{T}_L = \begin{pmatrix} 8.538840 & -9.248584 & 52.730020 \end{pmatrix}^T$, $\mathbf{T}_R = \begin{pmatrix} 9.072549 & -8.894637 & 334.634691 \end{pmatrix}^T$.

$$\mathbf{R}_L = \begin{pmatrix} -0.995351 & -0.002545 & -0.096282 \\ -0.003001 & 0.999985 & 0.004592 \\ 0.096269 & 0.004859 & -0.995344 \end{pmatrix}, \mathbf{R}_R = \begin{pmatrix} -0.981710 & -0.006470 & 0.190271 \\ -0.004249 & 0.999918 & 0.012080 \\ -0.190333 & 0.011051 & -0.981657 \end{pmatrix}.$$
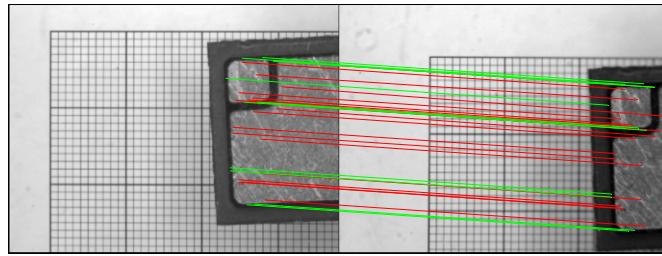
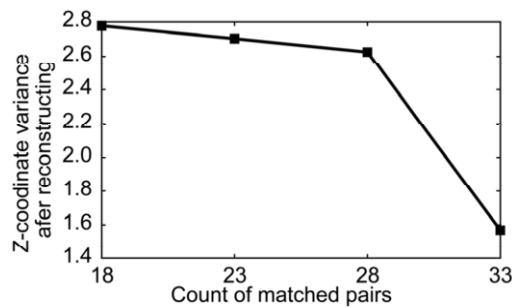Fig. 7. A schematic illustration of the stereo matching results.

Fig. 8. The relation between the quantity of matching elements and Z-axis variance after final reconstruction.

The SIFT features of the left and right views of IC chips were collected and shown in Figure 7. The red parts and green parts refer to matched feature pairs on the inner and outer side of the grooves, respectively. The total number of matched feature pairs was 33. In most cases, the height difference between the red pair plane and the green pair plane was measured. However, the measurement precision was limited by the insufficient quantity of feature point pairs. In this experiment, the quantity of feature point pairs was increased using a dynamic interpolation method, with appropriate matching elements as the benchmark. The experimental results revealed that the calculation variance decreases as the quantity, K, of matching elements increase. Therefore, the quantity of matching elements was maximized to reduce calculation errors. Figure 8 shows the unrectified variances on the Z-axis (after final reconstruction) obtained based on different amounts of matching elements.

The final results obtained based on the dynamic interpolation method and the depth rectification parameters are shown in Table 1. As shown, the average relative depth measured was 7.6738 μm. The deviations (compared with the designed values) are 26.637%, 16.733%, and 26.416%, respectively. As the platform moved from − 60μm to the origin, both the inner and outer depths increased; as the platform moved from the origin to + 60μm, the movements are significantly different. A possible reason for this discrepancy is the poor stability of the platform in the measurements. In other words, the gear clearance of the lead screw and platform vibration affected the uprising of the measuring position.

Another possible reason is the gap in the drive system of the machine, which may lead to deviations. For these reasons, the platform can be applied in measurements of object sizes. In other words, it can be used to measure the subject size and other relative geometric parameters, but not absolute location parameters.

4. Conclusions

Table 1

Final results obtained; units are μm

Position on the Z-axis	Inner depth of groove	Outer depth of groove	Groove depth
- 60	67.2645	59.9282	7.3363
0	127.5256	119.1989	8.3267
+ 60	129.6212	122.2628	7.3584

The stereo microscopic measurement is a key step in the processing of stereo signals. Stereo measurements of objects with few feature points were achieved using dynamic interpolation of feature point pairs based on stereo microscopic scaling. The SIFT feature matched pairs were obtained using conventional methods. The feature point pair interpolation was applied to a selected pair of matching elements, based on which the stereo reconstruction was conducted. In this way, the groove depths of IC chips were measured. Experimental results revealed that the effect of depth value on the depth measurement was insignificant, provided the interpolated feature points involved in the interpolation remained the same. Future studies may focus on effective interpolations of interpolated feature points and include the effects of radial and tangential distortions on microscopic measurements.

Acknowledgments

References

[1] G.W. Melenka, D.S. Nobes, J.P. Carey, et al., Three-dimensional deformation comparison of self-ligating brackets, American Journal of Orthodontics and Dentofacial Orthopedics **5** (2013), 645-657.

[2] Z.X. Hu, H.Y. Luo, Y.J. Du, et al., Fluorescent stereo microscopy for 3D surface profilometry and deformation mapping, Optics Express **10** (2013), 11808-11818.

[3] Z.W. Li and Y.F. Li, Dense and refined microstructure 3D measurement method based on an optical microscope and varying illuminations, Measurement Science and Technology **3** (2011), 35306-35317.

[4] K. Sharma, I. Moon and S.G. Kim, Extraction of visual landmarks using improved feature matching technique for stereo vision applications, IETE Technical Review **6** (2012), 473-481.

[5] K. Sharma, S.G. Kim and M.P. Singh, An improved feature matching technique for stereo vision applications with the use of self-organizing map, International Journal of Precision Engineering and Manufacturing **8** (2012), 1359-1368.

[6] J. Joglekar, S.S. Gedam and B.K. Mohan, Image matching using SIFT features and relaxation labeling technique-A constraint initializing method for dense stereo matching, IEEE Transactions on Geoscience and Remote Sensing **9** (2014), 5643-5652.

[7] H.C. Yang, M. Yu and S.B. Zhang, Wide baseline stereo matching based on scale invariant feature transformation with hybrid geometric constraints, IET Computer Vision **6** (2014), 611-619.

[8] C. Lee, J. Oh, C. Hong, et al., Automated generation of a digital elevation model over steep terrain in antarctica from high-resolution satellite imagery, IEEE Transactions on Geoscience and Remote Sensing **3** (2015), 1186-1194.

[9] C.Q. Zeng, J.F. Wang and P.J. Shi, A stereo image matching method to improve the DSM accuracy inside building boundaries, Canadian Journal of Remote Sensing **4** (2013), 308-317.

[10] Y.C. Wang, C.Y. Wu and P.C. Chung, Constraint-based correspondence matching for stereo-based interactive robotic massage machine, Journal of Intelligent and Robotic Systems **2** (2013), 179-196.

[11] M.L. Yang, Y.G. Liu, Z.S. You, et al., A homography transform based higher-order MRF model for stereo matching, Pattern Recognition Letters **4** (2014), 66-71.

[12] M. Antunes and J.P. Barreto, Symstereo: stereo matching using induced symmetry, International Journal of Computer Vision **3** (2014), 187-208.

[13] S. Kim, B. Ham, B. Kim, et al., Mahalanobis distance cross-correlation for illumination-invariant stereo matching, IEEE Transactions on Circuits and Systems for Video Technology **11** (2014), 1844-1859.

[14] V.D. Nguyen, D.D. Nguyen, S.J. Lee, et al., Local density encoding for robust stereo matching, IEEE Transactions on Circuits and Systems for Video Technology **12** (2014), 2049-2062.

[15] Y.G. Wang, G.Y. Jiang, M. Yu, et al., Calibration method of stereo microscope based on depth correction, Communications in Computer and Information Science **461** (2014), 109-120.

[16] D.G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision **2** (2004), 91-110.