Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Motion Stereo including Tracking Stability Analysis for 3-D Cable Reconstruction ⋆

Yukiyasu Domae * **Haruhisa Okuda** ** **Hidenori Takauji** *
Yuta Kimura * **Shun'ichi Kaneko** * **Takayuki Tanaka** *

* *Graduate School of Information Science and Technology, Hokkaido University, Kita 14, Nishi 9, Kita-ku, Sapporo, 060-0814, JAPAN (Tel: +81-(0)11-706-6761; e-mail: domae@ssc.ssi.ist.hokudai.ac.jp).*
** *Advanced Technology R&D Center, Mitsubishi Electric Corporation, 8-1-1, Tshukaguchi-Honmachi, Amagasaki, Hyogo 661-8661, JAPAN (e-mail: Okuda.Haruhisa@ct.MitsubishiElectric.co.jp)*

Abstract: We propose a novel approach to the three-dimensional reconstruction of flexible cables for applications in factory automation, such as cable handling and connecter insertion using robotic arms avoiding conflicts among multiple cables. The approach is based on motion stereo with a single vision sensor. To solve the stereo correspondence problem efficiently and effectively, feature-point projection with slit beams and a feature tracking algorithm based on a robust image-matching method is applied. In addition, we define the tracking stability of the feature points in order to reject defective stereo correspondences. Experiments using these methods demonstrate that an arched cable shape can be reconstructed to an accuracy of 1.5%.

Fig. 1. How to measure cable shapes.

1. INTRODUCTION

Automatic flexible cable handling is one of the principal topics in the field of factory automation. In this paper, a novel approach to the three-dimensional (3D) reconstruction of cables is proposed for the purpose of automatic handling. The method is intended for the reconstruction of the overall shape of stationary cables, and also as a preprocessing step for the extraction of more detailed parts, such as connecters.

Vision sensing is well suited to the acquisition of cable shape, being a non-contact method that is easy to install. One option is commonly used binocular vision. It is however difficult to solve the stereo correspondence problem when targeting flexible linear objects without rich textures. Indeed, even if multiple views are utilized, the problem remains far from easy. Light projection may be used to simplify the problems associated with cables when considering the effects of specular reflections. Laser range-finders, which are typical active sensing instruments, are susceptible to various effects and changes in illumination conditions. Complete cable surfaces cannot therefore be reconstructed stably using such active methods under real factory environment conditions. Visual cone intersection methods[1], and methods embodying the visual hull concept[2], while suitable for the reconstruction of complete cable shapes, require excessive space for installation. An ideal solution would be mountable on a robotic arm, cooperating with robot motion to accomplish tasks.

⋆ The authors gratefully acknowledge the contribution of the New Energy and Industrial Technology Development Organization (NEDO), and the reviewers' comments.

We deal with the problem using motion stereo with a single vision sensor mounted on a robotic arm, and discrete feature-point projection using slit beams. Motion stereo can handle a variable baseline length, tailored to suit a given task. Short baselines give rise to significant reconstruction errors because of the effects of stereo geometry and quantization noise. Long baselines instead make the stereo correspondence problem awkward to solve. Hence, variable baseline lengths tailored to tasks are advantageous. The discrete feature points in the slit beam projection provide a means to evaluate the reliability of the stereo correspondence, and perform reconstruction point by point. As stated above, reconstruction of an entire cable surface is not easy. One simplistic solution is projection in order to find the stable feature points on cables and reconstruct them separately. Assuming that a sufficient number of reliable feature points can be projected onto the cables, cable models based on the geometric constraints of cables may be fitted. One reconstruction method, the factorization method (Shape from Motion)[3] [4], may be suitable, except that all feature points under examination must be reliable in order to reconstruct objects, which is not feasible. Reconstruction on a point-by-point basis using motion stereo was therefore selected.

It is necessary to consider two problems in order to achieve this approach: the stereo correspondence problem of each projected feature point, and the evaluation of the reliability of each stereo correspondence. Multiple images taken from image sequences should be used in order to solve the first problem in a stable manner, as with multiple-baseline stereo[5]. The method of using line detection in closely sampled image sequences is stable but computationally expensive.[7][8] In order to achieve a stable solution with a low computational cost, we applied a tracking algorithm based on Orientation Code Matching (OCM)[6]. This image matching method is robust against changes of

Fig. 2. Feature point stereo correspondence.

illumination, and more sparsely sampled image sequences. The potential for ambiguity among the correspondences still remains, so long as the image features are subject to degradation caused by various factors, such as the quality of the projected feature points, surface gloss of cables, illumination conditions, backgrounds and so on. We define Tracking Stability (TS) by means of the epipolar constraint and statistical theory so as to limit the influence of the ambiguity. These methods facilitated stable and efficient 3D cable shape sensing.

In the next section, we outline the process involved in the reconstruction method. Two key algorithms, OCM and TS, are explained in Sections 3 and 4, respectively. In Section 5, experimental results regarding the reconstruction of arch-shaped cables are presented. Conclusions and future work are discussed in Section 6.

2. PROCESS OUTLINE

The outline of the process involved in our method is based on 3D reconstruction of discrete feature points using motion stereo, as mentioned above. Suppose that there are n feature points on the cables. The 3D position of the j^{th} feature point on the cables is denoted by $\mathbf{X}^j$ $(j = 1, 2, \cdots, n)$. The 3D position of the j^{th} feature point being reconstructed using the images at the t^{th} and $(t+k)^{th}$ frames is denoted by $\hat{\mathbf{X}}^j_{t/t+k}$. The 3D motion of the camera from the t^{th} to the $(t+k)^{th}$ position is denoted by $[\mathbf{R}_{t/t+k}, \mathbf{t}_{t/t+k}]$. Figure 1 shows twelve feature points reconstructed using a motion $[\mathbf{R}_{t/t+k}, \mathbf{t}_{t/t+k}]$. Such a segment of successive camera motion is defined as a motion element, e_t.

The steps of the algorithm are divided into four sections: feature extraction, feature tracking, defective point rejection and feature point reconstruction.

2.1 Feature Extraction

2D feature points are extracted from the image considered to be the first frame of the motion element e_t. The 2D position of the j^{th} feature point at the t^{th} frame is denoted by $\mathbf{x}^j_t$. These feature points are extracted using the difference between an image with slit beams and an image without slit beams. The feature points are defined as landmarks. Then, in order to handle absent landmarks during the reconstruction of cable shape, interpolated points are generated using an Orientation Code (OC)[10] texture analysis method which is described in Section 3.

2.2 Feature Tracking

At the $(t+1)^{th}$ frame, the stereo correspondences of the extracted feature points are found using OCM and an epipolar constraint. The epipolar equation[9] is defined as:

$$\tilde{\mathbf{x}}^{j\mathrm{T}}_t(\mathbf{t}_{t/t+1} \times (\mathbf{R}_{t/t+1}\tilde{\mathbf{x}}^{j\mathrm{T}}_{t+1} + \mathbf{t}_{t/t+1})) = 0 \qquad (1)$$

where $\tilde{\mathbf{x}}$ is the augmented vector of $\mathbf{x}$, and $\mathbf{x}^{\mathrm{T}}$ is the transposed vector of $\mathbf{x}$. The geometry of stereo correspondence is shown in Fig. 2. $\mathbf{C}_t$ is the position of a camera coordinate system at the t^{th} frame.

2.3 Defective Point Rejection

To determine and reject the defective points, we define the Tracking Stability (TS) of each feature point. A feature point that is being tracked well exists on the epipolar line, provided the camera calibration is accurate and the camera motion is acquired accurately. Assuming that feature points demonstrate uniform motion, a stably-tracked feature point is defined as a point with uniform motion on the epipolar line. TS, as described in Section 4, is based on the statistical nature of tracking errors and is used to find defective stereo correspondences. Unstable feature points can thus be rejected.

2.4 Reconstruction

$\hat{\mathbf{X}}^j$ is reconstructed by means of motion stereo applied to the 2D position of feature points with stable correspondence, $\mathbf{x}^j_t$ and $\mathbf{x}^j_{t+k}$, and the camera motion $[\mathbf{R}_{t/t+k}, \mathbf{t}_{t/t+k}]$. These reconstructed feature points are interpolated by cubic spline curves.

3. ORIENTATION CODE (OC) AND ITS APPLICATION

3.1 Definition

OC is defined as a quantized value corresponding to the orientation of the maximum intensity change in neighboring pixels[6]. The quantization level can be set arbitrarily, providing it is stable in the presence of intensity or contrast changes, highlights and shadows. The code c_{xy} is defined at an arbitrary pixel position as:

$$c_{xy} = \begin{cases} [\dfrac{\theta_{xy}}{\Delta_\theta}] & \text{if } |\bigtriangledown I_x| + |\bigtriangledown I_y| > \Gamma \\ N = \dfrac{2\pi}{\Delta_\theta} & \text{otherwise} \end{cases} \qquad (2)$$

Let θ_{xy}, Δ_θ and $\bigtriangledown I_i$ represent the direction of the gradient, the quantization width, and the derivative of I in the i-direction, respectively. Γ is a threshold for suppressing unreliable codes generated from neighbors with low contrast, for which an exceptional value of N is used. We adopted $N = 16$ throughout the paper.

3.2 OC Texture Analysis

Texture analysis using the entropy of the OC distribution in a local area is applied to find the feature points well suited to tracking. OC entropy[11] is defined as:

$$E_{xy} = -\sum_{i=0}^{N-1} \frac{h_{xy}(i)}{M^2 - h_{xy}(N)} log_2(\frac{h_{xy}(i)}{M^2 - h_{xy}(N)}) \qquad (3)$$

where $h_{xy}(i)$ $(i = 0, 1, \cdots, N)$ is defined as the number of the i^{th} OC in an M by M local region centered on any given pixel. OC texture analysis is robust against gradation and illumination change. It is not so much an edge [12] [13] or corner [14] [15] detector, as a rich texture detector containing corner detection.

Fig. 3. Hand-eye motion model based on motion elements.

Fig. 4. Examples of stable and unstable feature tracking.

Fig. 5. Description of j^{th} feature point motion between the i^{th} and $(i+1)^{th}$ frames in e_t.

Fig. 6. Standard normal distribution fitting results.

3.3 Orientation Code Matching (OCM)

Feature points are tracked using OCM[16], which can find templates under hostile conditions, such as illumination change, shading, highlighting, some amount of object deformation, and so on. As a similarity measure between a reference image f and a target image g of the same size, $L \times L$, the mean of the absolute residuals in OC, D, is defined as:

$$D = \frac{1}{M^2} \sum_{M \times M} d(c_f, c_g) \tag{4}$$

$$d(a,b) = \begin{cases} \dfrac{N}{4} & \text{if } a = N, b = N \\ min\{|a-b|, N - |a-b|\} & \text{otherwise} \end{cases} \tag{5}$$

4. TRACKING STABILITY (TS)

Correspondence among the feature points in the images is achieved using OCM between successive frames. However, because of the shapes of cables, poor textures caused by cables with specular surfaces, environmental effects and so on, complete correspondences are hard to retain in the image sequence. The TS of the feature points on the cables was thus defined in order to reject feature points which have large tracking errors.

4.1 Movement Model

Firstly, we define two constraint conditions for hand-eye motions: epipolar constraints and uniform motion of each feature point on an extended image plane. The extended image plane is defined as a plane including all the image planes in a motion element e_t. In order to satisfy these conditions, the motion of the robotic arm in a given motion element is expressed by four DoF affine transforms that contain translational motions along three axes and a rotation of the optical axis of the camera lens. The hand-eye movement model is shown in Fig. 3. When the camera takes samples at even intervals in a motion element, the two conditions above are assumed. Hand-eye systems can perform motions of sufficient complexity to achieve tasks by combining motion elements.

Under the movement model, the stably tracked feature points perform almost uniform motion on the epipolar lines. Some feature points however, move unsteadily on the epipolar lines as shown in Fig. 4. We define the TS of j^{th} feature point by using the variance of the movement distance $\mathbf{L}_i^j$, and the angle θ_i^j formed between the epipolar line with the direction of motion between the i^{th} and $(i+1)^{th}$ frames, as shown in Fig. 5. Formally, $\mathbf{L}_i^j$ and θ_i^j are defined as:

$$\mathbf{L}_i^j = \mathbf{x}_{i+1}^j - \mathbf{x}_i^j \tag{6}$$

$$\theta_i^j = \mathrm{Cos}^{-1}\left(\frac{\mathbf{L}_i^j \cdot \mathbf{e}_i^j}{|\mathbf{x}_{i+1}^j - \mathbf{x}_i^j|}\right) \tag{7}$$

where $\mathbf{e}_i^j$ is a unit tangent vector along the epipolar line.

4.2 Formulation based on The Statistical Nature of The Error

The sample variance of the Euclidean movement distance $L_i^j = |\mathbf{x}_{i+1}^j - \mathbf{x}_i^j|$ and the angle θ^j is defined as:

$$\begin{cases} {s_L^j}^2 = \dfrac{1}{k-1} \displaystyle\sum_{i=1}^{k-1} (L_i^j - \hat{L}^j)^2 \\ {s_\theta^j}^2 = \dfrac{1}{k-1} \displaystyle\sum_{i=1}^{k-1} (\theta_i^j - \hat{\theta}^j)^2 \end{cases} \tag{8}$$

where k is the number of the last frame in the motion element. Since the population parameter of the expected value $\hat{L}^j$ is not easily comprehended, the unknown parameter is chosen to be $\frac{1}{k}\sum_{i=1}^{k} L_i^j (= \bar{L}^j)$. In contrast, a population parameter of $\hat{\theta}^j$ may be found from our preliminary analysis, so $\hat{\theta}^j \equiv 0$.

Let us now consider the statistical nature of the errors, $L^j - \hat{L}^j$ and θ^j. Figure 6 shows histograms of the distance and angle errors observed in our experiments, fitted with the standard normal distribution. The experimental conditions are described in Section 5. The results reveal some fluctuation, but it is reasonable to assume that the errors obey normal-distributions, which we take to be $N(\bar{L}^j, {s_L^j}^2)$ and $N(0, {s_\theta^j}^2)$. When random variables $X_1, X_2, \ldots, X_n$ follow a normal distribution $N(\mu, \sigma^2)$, the following equation holds as a general characterization:

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1) \tag{9}$$

where $\chi^2(n-1)$, s^2 and σ^2 are an $n-1$ DoF chi-square distribution, the sample variance, and population variance, respectively. We therefore obtain:

$$\begin{cases} {s_L^j}^2 \sim \dfrac{\sigma_L^2}{k-2} \chi^2(k-2) \\ {s_\theta^j}^2 \sim \dfrac{\sigma_\theta^2}{k-1} \chi^2(k-1) \end{cases} \tag{10}$$

The DoF of the chi-square distribution of L becomes $k-2$, because the population parameter of $\hat{L}^j$ is unknown.

We define thresholds of ${s_L^j}^2$ and ${s_\theta^j}^2$ using the variance values corresponding to a 100(1-α) percent confidence interval for the chi-square distribution:

$$\begin{cases} \Gamma_L = \dfrac{\sigma_L^2}{k-2} \chi_\alpha^2(k-2) \\ \Gamma_\theta = \dfrac{\sigma_\theta^2}{k-1} \chi_\alpha^2(k-1) \end{cases} \tag{11}$$

Fig. 7. Environment used in the cable measurement experiment.

Fig. 8. An example cable image sequence.

Fig. 9. Landmarks on cables using slit beam irradiation.

Fig. 10. 40 feature points extracted from cables.

Fig. 11. SCC rate in each frame with or without tracking.

We adopted $\alpha = 0.01$ throughout the paper.

Because σ_L^2 and σ_L^2 are in general unknown, appropriate values must be estimated. In our experiments, these values were estimated from the variance among all feature points. Finally, the TS is defined as:

$$T^j = T_L^j T_\theta^j \tag{12}$$

$$T_L^j = \begin{cases} 1 - \dfrac{{s_L^j}^2}{\Gamma_L} & \text{if } {s_L^j}^2 \leq \Gamma_L \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

$$T_\theta^j = \begin{cases} 1 - \dfrac{{s_\theta^j}^2}{\Gamma_\theta} & \text{if } {s_\theta^j}^2 \leq \Gamma_\theta \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

When the observed variance is outside of a 99 percent confidence interval for the chi-square distribution, T_L and T_θ become zero. If the variances ${s_L^j}^2$ and ${s_\theta^j}^2$ are small, T_L and T_θ approach one.

5. MEASUREMENT EXPERIMENT

5.1 Experimental Environment

The experimental environment is shown in Fig. 7. A camera was mounted on an automatic X-Z stage with a positioning accuracy of 0.02 mm. The depth from the floor to the CCD plane of the camera was 450 mm. The camera was moved 140 mm along the X_I-axis of the image in 7 mm intervals, and 20 images were taken at each position. An example image sequence is shown in Fig. 8. Motion stereo was applied using the 1^{st} and 20^{th} images. This is equivalent to stereo vision with a baseline of 140 mm. The images were of VGA size, and the maximum frame rate was 30 fps. The focal length of the lens was 6 mm. The intrinsic camera parameters were computed using the flexible estimation method by Zhang[17]. The parameters were then modified by measuring the 3D positions of the corners of a checker board to an accuracy of $1/10^{th}$ pixel. By using this measurement system, we calculated the depth of the corners of the checker board to within 0.5% of their true values. This illustrates the basic performance of this 3D measurement system.

Fig. 12. TS of each feature point.

Fig. 13. 3D cable shape reconstruction.

5.2 Feature Extraction Performance

20 landmarks and interpolated points were extracted from two arch-shaped cables as shown in Fig. 10. The interpolated points were extracted from the cable surface in a stable manner. Many of the landmarks were on the upper side, due to the angle of rotation. In addition, a number of landmarks, *e.g.*, 6, 15, and 18, were not on the cables, because reflections of some laser slit beams projected onto the background were seen on the cables in the image. These are false extractions, but a correct choice with respect to the image. 82.5% of the feature points were on the cables.

5.3 Feature Tracking Performance

Firstly, we have tested the effectiveness of the tracking algorithm by comparison to stereo correspondence without tracking. "Without tracking" means solving the stereo correspondence problem using only two images for each frame, the first and the last. The results are shown in Fig. 11. The Stereo Correct Correspondence (SCC) rate, shown in Fig. 11, is defined as SCC= $100\frac{n-n_d}{n}$%, where n is the number of feature points and n_d is the number of falsely tracked feature points. False tracking was defined as tracking with an error of more than 5 pixels. The error was observed by hand in this experiment. The results with tracking are improved, and have an increased number of frames. About 40% of the feature points in the last frame are however tracked falsely.

Figure 12 shows the TS of each feature point in the last frame. The TS of the landmarks has an average value of 0.79, while the TS of the interpolated points has an average value of 0.53. Five interpolated points, 24, 27, 37, 38 and 39, were rejected according to the threshold based on the confidence interval of the chi-square distribution. Without these points, the average for the interpolated points rises to 0.71. The TS of the falsely extracted landmarks was lower than the correctly extracted landmarks, as mentioned above, because they move in a different direction to the camera's motion in the extended image plane. In addition, some landmarks, such as 11, disappeared during the tracking process because of specular reflections on the surface of the cable from lights in the environment. The TS of such feature points was lower.

5.4 3D Reconstruction Performance

Figure 13 (a) shows the 3D reconstruction results for all feature points and the application to the cubic spline interpolation. The errors in stereo correspondence cause a noisy reconstruction result. On the other hand, in Fig. 13 (b), the reconstruction using only points with high TS ($T > 0.8$) reveals results with a better appearance than when all points are used.

For a quantitative evaluation, a non-contact 3D digitizer VIVID 910 from KONIKA MINOLTA was used to reconstruct the arch-shaped cables. The digitizer was not

Table 1. Height measurement results for arch-shaped cables.

	Digitizer	High TS	All points
cable A	105mm	110mm (+5)	116mm (+11)
cable B	110mm	114mm (+4)	116mm (+6)

optimized for 3D cable reconstruction. The experiment was therefore performed in a dark room with no environmental lighting. The positioning accuracy of the digitizer was 0.008 mm. The heights from the floor to the arch-shaped cables were measured using the digitizer as well as by the proposed method. The error between the result of the digitizer and the proposed method for each reconstructed point was measured. Next, the maximum error was found. Table 1 shows the results for the reconstruction of the heights of each cable top. The reconstruction results for cable B are better than for cable A, because many interpolated points, 37, 38, 39 and so on, whose TS values were zero appeared on cable B.

The result using only points with high TS values is closer than the result using all the points. The depth errors in each reconstructed point when using all the points are within 3.2%. By contrast, the errors when using only points with a high TS value are within 1.5%. Errors were cut by over 50% by means of TS analysis.

6. CONCLUSIONS

We proposed a system for reconstructing the shapes of cables in 3D, using a single camera intended for mounting on a robotic arm. The method involves landmark projection, OC-based feature tracking, TS analysis and motion stereo. In the arch-shaped cable reconstruction experiments, cable shape was reconstructed to within 1.5% of its true depth value. Future work includes the improvement of the landmark projection method.

REFERENCES

[1] K.N. Kutulakos and S.M. Seitz, "A theory of shape by space carving," Proc. of IEEE Conf. on Computer Vision, pp.307-314, 1999.

[2] A. Laurentini, "The visual hull concept for silhouette-based image understanding," IEEE Trans. of Pattern Analysis and Machine Intelligence, 16(2), pp.150-162, 1994.

[3] C. Tomashi, T. Kanade, "Shape and Motion from Image Streams under Orthography: a Factorization Method," Journal of Computer Vision, 9:2, pp.137-154, 1992.

[4] C. J. Poelman, T. Kanade, "A Paraperspective Factorization Method for Shape and Motion Recover," IEEE Trans. of Pattern Analysis and Machine Intelligence, vol.19, no.3, pp.206-218, 1997.

[5] M. Okutomi, T. Kanade, "A multiple-baseline stereo method," Proc. of CVPR, pp.63-69, 1991.

[6] F. Ullah, S. Kaneko and S. Igarashi "Orientation code matching for robust object search," IEICE Trans. of Inf. & Sys, E84-D(8), pp.999-1006, 2001.

[7] M. Yamamoto, "The Image Sequence Analysis of Three-Dimensional Dynamic Scenes," Trans. of Institute of Electrical Engineers of Japan, vol.J69-D, no.11, pp.1631-1638, 1986. (in JAPANESE)

[8] D. Jelinek and C.J. Taylor "Quasi-Dense Motion Stereo for 3D View Morphing" Int. Symp. on Virtual and Augmented Architecture (VAA01) pp 219-229, June 2001.

[9] H.C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections," Nature, 293, pp.133-135, 1981.

[10] F. Ullah, S. Kaneko and S. Igarashi, "Tracking looming and receding objects by orientation code matching," Journal of IIEEJ, vol.31. No. 1, pp.94-102, 2002.

[11] H. Takauji, S. Kaneko and T. Tanaka, "Robust Tagging in Strange Circumstance," Electrical Engineering in Japan, vol.156, no.4, pp.22-32, 2006.

[12] I. Sobel, "Neighbourhood coding of binary images for fast contour following and general array binary processing," Computer Graphics and Image Processing, vol.8, pp.127-135, 1978.

[13] J. Canny, "A Computational Approach to Edge Detection", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.8. No. 6, 1986.

[14] H. Moravec, "Towards Automatic Visual Obstacle Avoidance," Proc. 5th Int. Joint Conf. Art. Intell, pp.584, 1977.

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

[16] Y. Domae, S. Kaneko and T. Tanaka, "Robust Tracking based on Orientation Code Matching under Irregular Conditions," Proc. of SPIE, Vol.6051, 0S, 2005.

[17] Z. Zhang, "A Flexible New Technique for Camera Calibration," IEEE Trans. on Pattern Analysis and Machine Intelligence, 22(11):1330-1334, 2000.