An Iterated Optical Flow Estimation Method for Automatically Tracking and Positioning Homologous Points in Video Image Sequences

Associate Prof. Dr.-Ing. Jaan-Rong Tsay1
Department of Geomatics, National Cheng Kung University
No. 1, University Road, Tainan 701, Taiwan, R.O.C.
tsayjr@mail.ncku.edu.tw

M. Sc. I-Chien Lee2
EverTime Technology, Inc.
3F, No. 107, Sec. 2, Roosevelt Rd., Taipei 106, Taiwan, R.O.C.
john@evertime.com.tw

Abstract: The optical flow theory can be utilized for automatically tracking and positioning homologous points in digital video (DV) image sequences. In this paper, the Lucas-Kanade optical flow estimation (LKOFE) method and the normalized cross-correlation (NCC) method are compared and analyzed using the DV image sequences acquired by our SONY DCR-PC115 DV camera. Thus, an improved optical flow estimation procedure, called "Iterated Optical Flow Estimation (IOFE)", is presented. Our test results show that the trackable range of 3~4 pixels in the LKOFE procedure can be apparently enlarged to 30 pixels in the IOFE.

Keywords: Optical Flow, Digital Video Image Sequence.

1. Introduction

Automatic real-time mobile mapping vehicle system (MMVS) is being developed e.g. in the field of geomatics for the real-time surveying and mapping purposes. In that system, automatic DV image point extraction, point transfer, and point localization are useful, and occasionally necessary, techniques, where point transfer on terrestrial and close-range images can be done by automatic point tracking. In the current related point tracking techniques, [1] incorporates the epipolar geometry into the random sample consensus (RANSAC) method proposed in [2] for efficient and reliable wide baseline image matching. Nevertheless, automatic wide baseline image matching techniques still need to be further studied [3,4]. A least squares matching tracking algorithm was also proposed for human body modeling in [5]. The normalized cross correlation (NCC) method was used for image tracking and positioning targets at sea by using video images taken on a helicopter in [6]. The Kalman filter was applied in [7] to predict matching areas and thus to decrease the size of searching window of correlation-based tracking. [8] also adopted the Kalman filter to develop a new method for occlusion robust adaptive template tracking.

Gradient-based optical flow estimation approaches are simple for tracking image sequences [10]. In 1981, two famous approaches are proposed: the LKOFE and the Korn-Schunck optical flow (H-S) method. The LKOFE computes most easily and fast. It is often used in the devices of high image frame rate. The H-S method is suitable for the image vector field with continuous smooth variation. It must be computed iteratively. Under the circumstances, the LKOFE method will be adopted in our approach.

In this paper, a newly developed method, called "iterated optical flow estimation (IOFEA)" approach, is used to improve the function of traditional optical flow method. Also, both optical flow method and normalized cross-correlation (NCC) method are to be analyzed and compared concerning with the effectiveness and quality.

2. The IOFE Approach

Fig. 1 shows the automatic point transfer algorithm for DV images. The originally acquired DV signal is converted into sequential DV images by the free software TMPGEnc available at http://www.tmpgenc.com. Now, the DV images as shown in Fig. 2 are interlaced. They must be de-interlaced. Blurred de-interlaced images will be automatically selected by simply using average image gradient. Blurred images will not be used. Then, feature points are extracted by using the Förstner operator [11]. Those feature points "flowing" into a homogeneous area will be deleted. The LKOFE, NCC, and the proposed IOFE methods are then used for point tracking. In the final step, tracking errors are detected by using the least squares adjustment and correlation coefficient check.

This paper uses the LKOFE method to build the optical flow vector model, and the finite differences approach and the block motion model to estimate the related gradients at a point for a image function $S(x,y,t)$ dependent on positional and time variables x, y, and t. Detailed formulas can be found in [12]. Thus, a displacement vector at a pixel P can be computed typically from a template mask and a searching mask of the same size centered at the (r,c)-th pixel in the sequential two DV images, respectively. Compared with the typical LKOFE method, our IOFE method changes this rule and involves the following steps:

1. Compute the displacement vector (dr, dc) from a template and a searching mask of m×m pixels (e.g. m=11)

centered at the (r,c)-th pixel P in the sequential two DV images.

2. The template mask remains the same. Move the searching mask from (r,c) to (r+dr, c+dc). Compute again a new displacement vector (dr', dc').
3. If dr'≠0 or dc'≠0, repeat the step 2. Otherwise, stop the computation at the pixel P.

Normally, the computation is completed after 2 or 3 iterations, if the displacement vector length is less than 3 pixels. If the number of iterations is larger than 10, stop the divergent computation and label the pixel P as an "invalid point". Otherwise, label the pixel P as an "valid point".

DV image acquisition and preprocessing
- DV image acquisition
- Convert DV signal into sequential DV images
- De-interlace DV images
- Blurred image selection and deletion

Feature point extraction and selection
- Feature point extraction by the Förstner operator
- Feature point selection

Feature point tracking
- Point tracking by LKOFE, IOFE, NCC
- Error detection

Fig. 1. Flowchart of data processing in the automatic point transfer (APT) algorithm used in this paper.

Fig. 2. Interlaced DV images.

3. Test results

Fig. 3 shows a DV image of near 2D objects on a wall. It sequential images are used as test images. Fig. 4 shows the histograms of displacement vector lengths at all valid points for tracking from the 1st image to 2nd~7th image, respectively. It expresses clearly that the number of valid points (or trackable points) is decreased, if the time interval of the aforementioned image function $S(x,y,t)$ is increased. The number of trackable points is 32% at one image interval, and is continuously reduced to 0% at the time interval of 8 images (from image 1 to 9). Moreover, a second top wave curve emerges in the histogram curves (C)~(F). It means that a large number of points with a displacement vector length of 18~38 pixels still are trackable. Furthermore, these histograms also show that a large number of points (78%~99%) are wrong tracked, since their displacement vector lengths are less than the related image shift distance. Therefore, a mechanism for error detection on the tracking results is necessary. As shown in Fig. 5, the LKOFE method determines a large number of points with shorter displacement vectors than the real ones. Its registration accuracy is 1 pixel, where the affine trasnfoprmation is used as the registration model. The IOFE method has a registration accuracy of 0.511 and 0.415 pixel, respectively, if error-deletion is not or is done. Fig. 6 shows that the IOFE method generates tracked point pairs with higher correlation than the LKOFE. Table 1 shows the statistic figures of this set of test images. It shows that the NCC method has the best registration accuracy and provides most valid points, but is most time-consuming.

The same DV images are also used for the tests with different mask size. The results show that the maximal trackable range almost remains the same, although the mask size is increased from 11×11 to 41×41.

Fig. 7 shows another test results. Visual check verifies that the IOFE method provides better results than the LKOFE method.

Fig. 8 and Table 2 show that both IOFE and NCC method can efficiently track points for DV images of 60 fps (=frames per second).

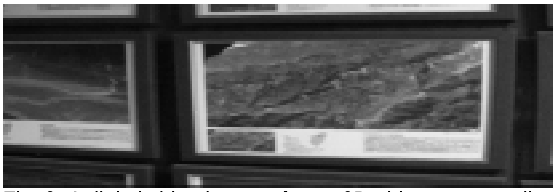

Fig. 3. A digital video image of near 2D objects on a wall.

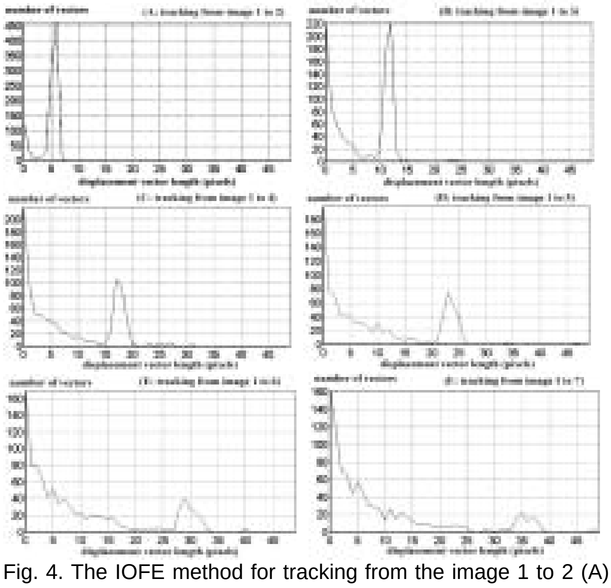

Fig. 4. The IOFE method for tracking from the image 1 to 2 (A), from 1 to 3(B), from 1 to 4 (C), from 1 to 5 (D), from 1 to 6 (E), and from 1 to 7 (F) (number of extracted feature points = 650).

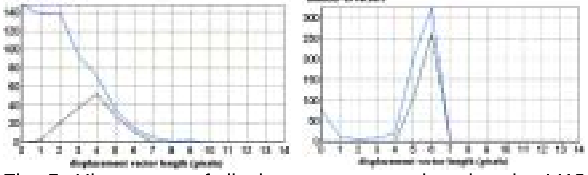

Fig. 5. Histograms of displacement vector lengths: the LKOFE method (left) and the IOFE method (right), where light-blue and dark-blue lines denote the ones without and with error detection.

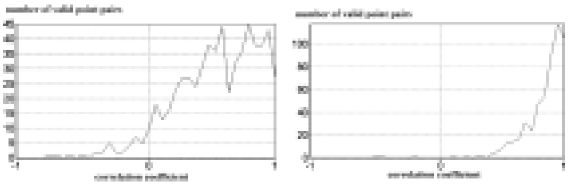

Fig. 6. Curves of correlation coefficients on all tracked point pairs by the LKOFE (left) and the IOFE method (right).

Table 1. Statistic figures of tracking results by three methods (number of extracted feature points = 650).

	LKOFE	IOFE	NCC
number of valid points	115 (17%)	312 (48%)	366 (56%)
registration accuracy (pixels)	0.796	0.415	0.351
RMSD X: Y:	 0.685 0.384	 0.306 0.277	 0.310 0.162
Computation time (second)	0.55	0.59	5.47

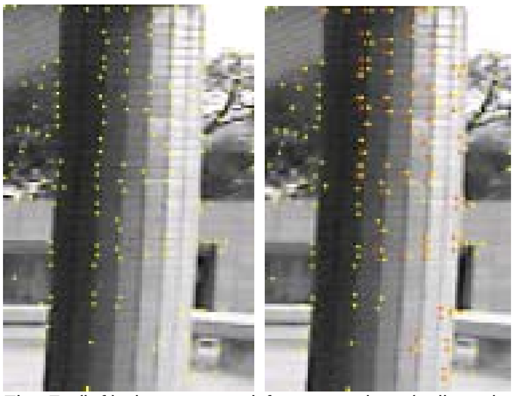

Fig. 7. (left) the extracted feature points (yellow dots) overlaid with DV image; (right) the tracked points by the LKOFE (red dots) and by the IOFE (yellow), where each point with the same tracking results is denoted by a single yellow point.

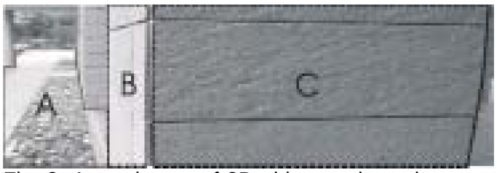

Fig. 8. A test image of 3D objects, where three sets of registration parameters are used for the areas A, B, and C, respectively.

Table 2. Statistic figures for DV images shown in Fig. 8

		IOFE	NCC
A	Number of valid points	21%	39%
	$\hat{\sigma}_0$ (pixels)	0.140	0.000
B	Number of valid points	12%	12%
	$\hat{\sigma}_0$ (pixels)	0.149	0.154
C	Number of valid points	22%	24%
	$\hat{\sigma}_0$ (pixels)	0.256	0.275

4. Conclusions

This paper presents a new approach for automatic point tracking in sequential digital video (DV) images, called "iterated optical flow estimation (IOFE)". Compared with the traditional LKOFE method, the IOFE approach significantly increases the maximum tracking distance and also improves the reliability of the tracking results. Test results show that the tracking range is increased from 3-4 pixels in the LKOFE method to about 30 pixels in the IOFE approach. Moreover, the mechanism for error-detection (by the average gradient, NCC, and least squares adjustment) can efficiently detect and delete wrong tracked points and thus improve the quality of point tracking. However, some further studies are still needed, e.g. high precision point measurement with a sub-pixel accuracy level, and rules for adding new tracking points. Thus, the IOFA can be improved so that it can be utilized in photo triangulation for the MMVS.

Acknowledgement

We sincerely appreciate Prof. King-Chang Lo at the Department of Geomatics, National Cheng Kung University, Taiwan, R.O.C., for providing the SONY DCR-PC115 DV camera.

References

[1] Hartley, R., and Zisserman, A., 2000. *Multiple View Geometry in Computer Vision*. Cambridge University Press.

[2] Fischler, M.A., and Bolles, R.C., 1981. Random Sample Consensus: a Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography. *Commun. Assoc. Comp. Mach*., 24: 381-395.

[3] Pritchett, P., and Zisserman, A., 1998. Wide Baseline Stereo Matching, *6th International Conference on Computer Vision*, Bombay, India, pp. 754-760.

[4] Van Gool, L., Tuytelaars, T., Ferrari, V., Strecha, C., Vanden Wyngaerd, J., Vergauwen, M., 2002. 3D Modeling and Registration Under Wide Baseline Conditions, *IASPRS Commission III*, pp. A3-A14.

[5] D'Apuzzo, N., Plaenkers, R., and Fua, P., 2000. Least Squares Matching Tracking Algorithm for Human Body Modeling. *IAPRS* Vol. XXXIII, Part B5, Amsterdam, pp. 164-171.

[6] Chen, M.-H., and Chen, C.-F., 2002. Target Positioning Using Object Tracking Technique for Video Images, ACRS2002, pp. 1-6.

[7] Hong, S.-M., 2002. Steady-state Analysis of Computational Load in Correlation-based Image Tracking. IEE Proc.-Vis. Image Signal Processing, Vol. 149, No. 3.

[8] Nguyen, H.T., Worring, M., and Boomgaard, M., 2001. Occlusion Robust Adaptive Template Tracking. *8th IEEE International Conference on Computer Vision*, Vol. 1, pp. 678-683.

[9] Fuse, T., Shimizu, E., and Tsutsumi, M., 2000. A Comparative Study on Gradient-Based Approaches for Optical Flow Estimation. *IAPRS*, July 16-23, Vol. 33, Part B5/1, pp. 269-276.

[10] Lim, S., and El Gamal, A., 2001. Optical Flow Estimation Using High Frame Rate Sequence, *Proceedings 2001 International Conference on Image Processing*, Vol. 2, pp. 925-928.

[11] Förstner, W., 1993. Feature Extraction in Digital Photogrammetry, *Photogrammetry Record*, 14(82): 595-611.

[12] Tekalp, M., 1995. *Digital Video Processing*. Prentice Hall.