A NEW APPROACH FOR AUTOMATIC SELECTION AND TRANSFER OF CORRESPONDING IMAGE POINTS IN DIGITAL VIDEO IMAGE SEQUENCES

J.-R. Tsay [a], I.-C. Lee [b]

[a] Dept. of Geomatics, Natl. Cheng Kung University, No. 1, University Rd., 701 Tainan, Taiwan - tsayjr@mail.ncku.edu.tw
[b] EverTime Technology, Inc., 3F, No. 107, Sec. 2, Roosevelt Rd., 106 Taipei, john@evertime.com.tw

Commission III, IC WG III/5

KEY WORDS: Digital, Video, Close Range, Tracking, Registration, Algorithms

ABSTRACT:

We present a new approach for automatic selection and transfer of corresponding image points in digital video (DV) image sequences. This approach utilizes and extends some well-known techniques and theories (e.g. the optical flow theory) as well as a designed simple error-detection mechanism to achieve a much better efficiency than the well-known Lucas-Kanade optical flow estimation (LK) method. Our test results use the SONY DCR-PC115 DV image sequences and show that the trackable range of 3-4 pixels in the LK method can apparently be enlarged to 30 pixels in this new approach. The proposed error-detection mechanism simply utilizes the average gradient, normalized cross-correlation, and a simple image registration aided by least squares adjustment. Test results show that it can efficiently detect and delete wrong tracked points, and thus improve the quality of point transfer, and provide accurate coordinates of a large number of corresponding image points in DV image sequences. This work aims at high precision automatic image triangulation for the automatic real-time mobile mapping vehicle system (MMVS). Some future works need to be done.

1. INTRODUCTION

In the field of geomatics, automatic real-time mobile mapping vehicle system (MMVS) is being developed for the real-time digital-map-updating and other surveying purposes. In that system, those techniques for automatic image point extraction, point transfer, and point measurement are useful, and occasionally necessary. Therein, corresponding points in terrestrial or close-range stereo images can be found in a full- or semi-automatic manner by different image matching techniques or by differential methods (Trucco & Verri, 1998).

Image matching techniques may utilize low-level data such as raw image values, middle-level data such as edge features, or high-level data such as symbolic and topological relationships. Also, they are often integrated with other disciplines and applied in the so-called wide-baseline image matching and tracking. For example, a least squares matching tracking algorithm was also proposed for human body modeling in (D'Apuzzo et al., 2000). The normalized cross correlation (NCC) method was used for image tracking and positioning targets at sea by using DV images taken on a helicopter in (Chen & Chen, 2002). The Kalman filter was applied in (Hong, 2002) to predict matching areas and thus to decrease the size of searching window of correlation-based tracking. (Nguyen et al., 2001) also adopted the Kalman filter to develop a new method for occlusion robust adaptive template tracking. (Hartley & Zisserman, 2000) incorporates the epipolar geometry into the random sample consensus (RANSAC) method proposed in (Fischler & Bolles, 1981) for efficient and reliable wide-baseline image matching. Nevertheless, automatic wide-baseline image matching techniques still need to be further studied (Pritchett & Zisserman, 1998; Van Gool et al., 2002).

Differential methods such as the so-called optical flow approach are suitable for tracking corresponding points in a series of DV images, where two consecutive images have a very short baseline. Gradient-based optical flow estimation approaches are simple techniques for tracking image sequences (Lim & El Gamal, 2001). In 1981, two famous approaches are proposed: the LK method and the Horn-Schunck optical flow (H-S) method. The LK method computes most easily and fast, but it can track only those short image displacement vectors. It is often adopted in digital signal processors as well as in the devices of high image frame rate. The H-S method computes slowly and iteratively. It is suitable for the image displacement vector field with local continuous smooth variation. Under the circumstances, the LK method will be adopted and improved in our approach.

In this paper, a newly developed approach, called "iterated optical flow estimation (IOFE)" 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 compared concerning their effectiveness and quality.

2. THE NOVEL APPROACH

In order to simplify the problem to be solved, it is assumed that only those sequential DV images of immovable objects in a scene are to be taken by a DV camera in a moving mode. Thus, our approach is expected to provide photo coordinates of a large number of corresponding image points of immovable object points to photo triangulation application in a MMVS. Photo triangulation determines then 3D object coordinates of all pass points and orientation parameters of each image, where all pass points are assumed to be immovable and to have fixed 3D object coordinates. These data can be further used to reconstruct each stereomodel or to determine the 3D object coordinates of all interest features in a MMVS.

Figure 1 shows our automatic point extraction and transfer algorithm for DV images. The originally acquired DV signal is first converted into sequential DV images by the free software TMPGEnc available at http://www.tmpgenc.com. Now, the DV images as shown in Figure 2 are interlaced. They must be de-interlaced. Blurred de-interlaced images will be automatically selected by simply using average image gradient. A de-interlaced image will be tagged as a blurred image, if it has a small average image gradient less than 10% of the average gradient of five nonblurred de-interlaced images before it. Blurred images will not be used. Then, feature points are extracted by using the Förstner operator (Förstner, 1993) because it can extract as clear and definite features as possible, such as corner points. Those feature points "flowing" into a homogeneous area will be deleted. The LK, NCC, and our 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.

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 LK, IOFE, NCC
- Error detection

Figure 1. Flowchart of data processing in the automatic point extraction and transfer (APET) algorithm used in this paper

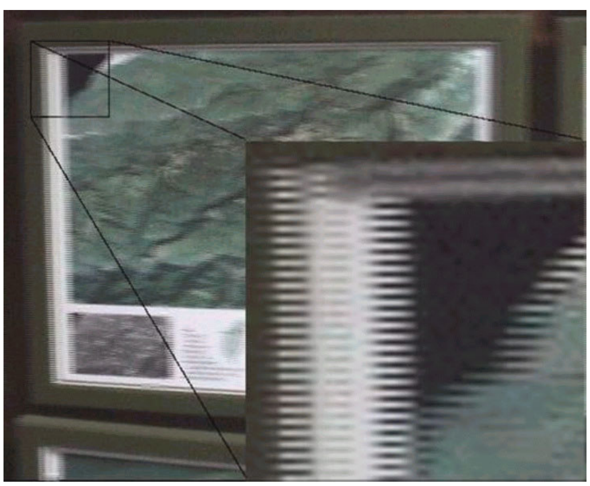


Figure 2. Interlaced DV images

This paper uses the LK method to build the optical flow vector model, and utilizes the finite difference 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 (Tekalp, 1995). Thus, a displacement vector at a pixel P can be computed typically from a template mask and a searching mask of the same size with its centre at the (r,c)-th pixel in two sequential DV images, respectively. Compared with the typical LK 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) with its centre at the (r,c)-th pixel P in two sequential 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 a "valid point".

3. EXPERIMENTS AND ANALYSES

Figure 3 shows a DV image of near 2D objects on a wall. Their sequential DV images are used as test images. Figure 4 shows the histograms of displacement vector lengths at all valid points for tracking from the 1st image to 2nd~7th image, respectively. It illustrates 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). Nevertheless, 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. Also, these histograms 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 Figure 5, the LK method determines a large number of points with shorter displacement vectors than the real ones. Its registration accuracy is 1 pixel, where the affine transformation 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. Figure 6 shows that the IOFE method generates tracked point pairs with higher correlation than the LK method. 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. Figure 7 shows some test results of a 3D scene. Visual check verifies that the IOFE method provides better results than the LK method. Figure 8 and Table 2 show that both IOFE and NCC method can efficiently track points for DV images of 60 fps (=frames per second).

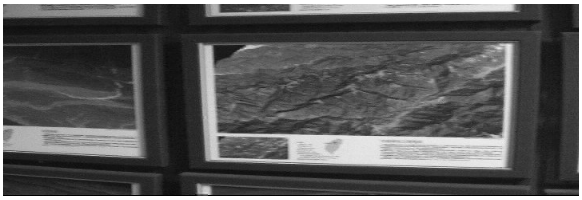
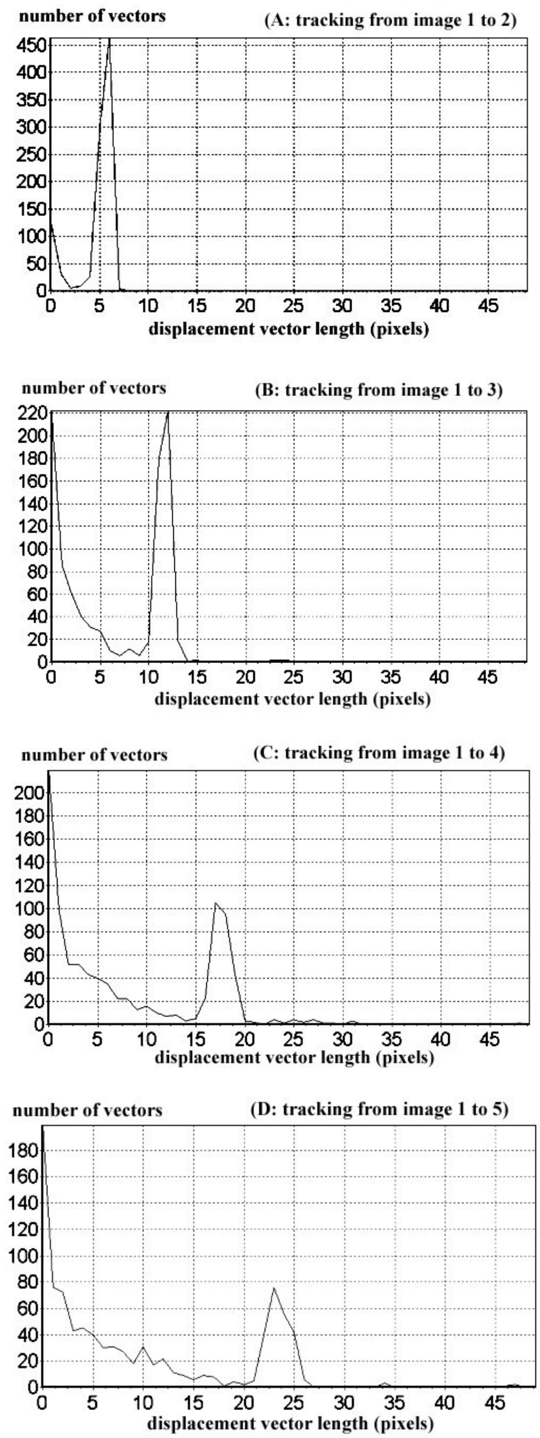


Figure 3. A digital video image of near 2D objects on a wall



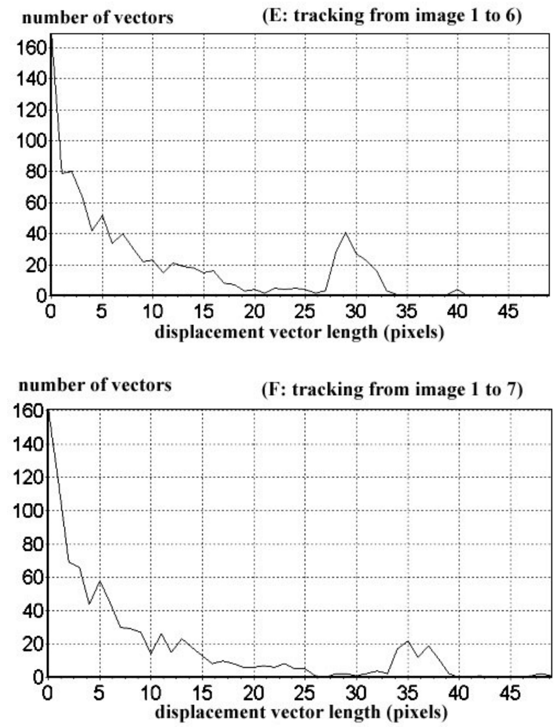


Figure 4. The IOFE method for tracking from the image 1 to 2 (A), from 1 to 3(B), from 1 to 4 (C), from 1 to 5 (D), from 1 to 6 (E), and from 1 to 7 (F) (number of extracted feature points = 650)

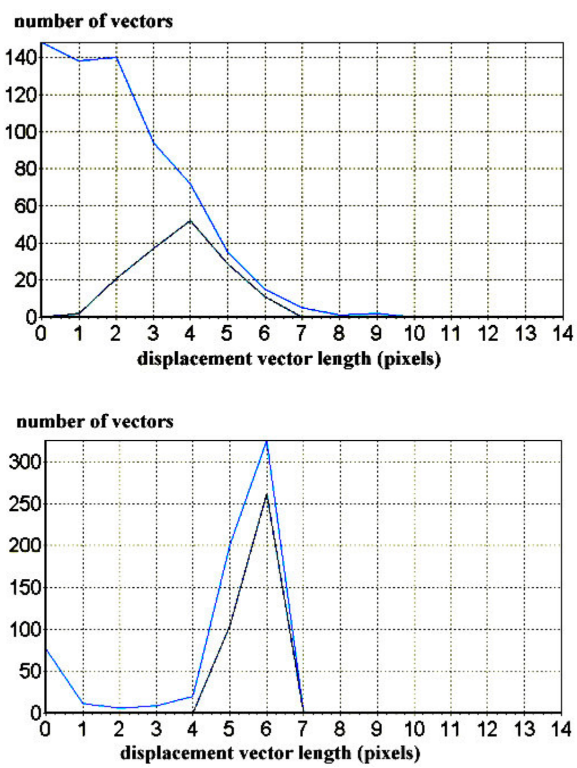


Figure 5. Histograms of displacement vector lengths: the LK method (top) and the IOFE method (bottom), where light-blue and dark-blue lines denote the ones without and with error detection

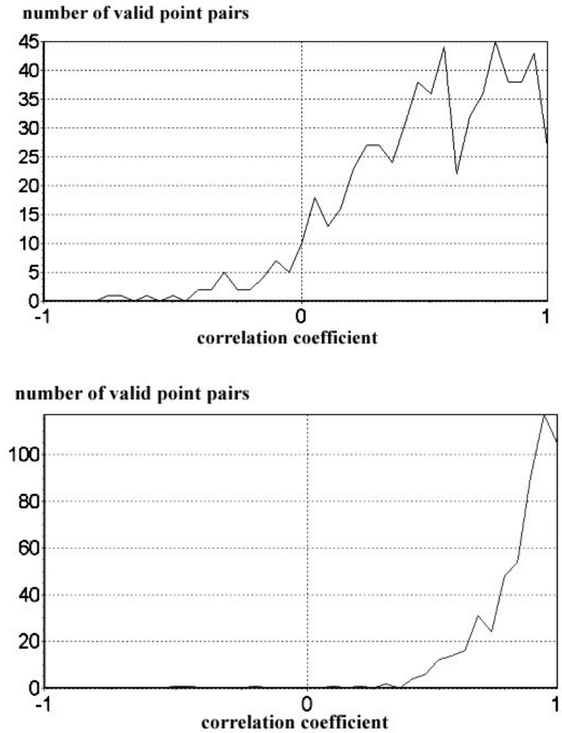


Figure 6. Curves of correlation coefficients on all tracked point pairs by the LK method (top) and our IOFE method (bottom)

	LK	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

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

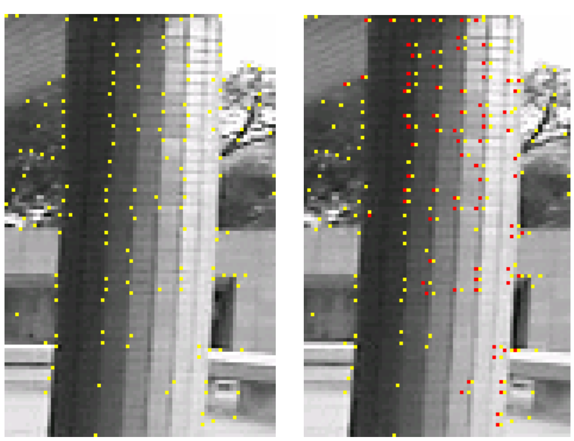


Figure 7. (left) the extracted feature points (yellow dots) overlaid with DV image; (right) the tracked points by the LK (red dots) and by the IOFE (yellow), where each point with the same tracking results is denoted by a single yellow point

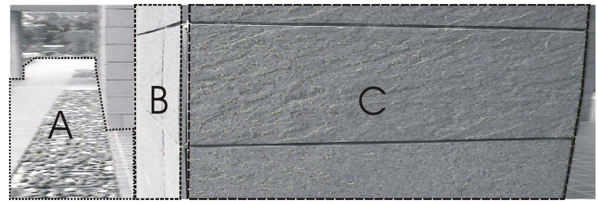


Figure 8. A test image of 3D objects, where three sets of registration parameters are used for the areas A, B, and C, respectively

		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

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

4. CONCLUSIONS

In this paper, we present a new approach for automatic extraction, selection and transfer of corresponding image points in a series of sequential digital video (DV) images. It is called “iterated optical flow estimation (IOFE)”. This approach utilizes and extends some well-known techniques and theories (e.g. the optical flow theory) as well as a proposed simple error-detection mechanism to achieve a much better efficiency than the well-known Lucas-Kanade optical flow estimation (LK) method. Compared with the traditional LK method, the IOFE approach significantly increases the maximum tracking distance and also improves the reliability of the tracking results. Our test results use the SONY DCR-PC115 DV image sequences and show that the trackable range of 3-4 pixels in the LK method can apparently be enlarged to 30 pixels in this new approach.

Moreover, the proposed error-detection mechanism simply utilizes the average gradient, normalized cross-correlation, and a simple image registration aided by least squares adjustment. Test results show that it can efficiently detect and delete wrong tracked points, and thus apparently improve the quality of automatic point transfer, and automatically provide accurate coordinates of a large number of corresponding image points in DV image sequences.

This work aims at high precision automatic image triangulation for the automatic real-time mobile mapping vehicle system (MMVS). Some future works need to be done, e.g. high precision point measurement with a sub-pixel accuracy level, and rules for adding new tracking points. Thus, the IOFE might be improved so that it can be utilized in high precision photo triangulation for the real-time map-updating and other surveying purposes of a MMVS.

REFERENCES

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

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

Fischler, M.A., and Bolles, R.C., 1981. Random Sample Conse nsus: a Paradigm for Model Fitting with Application to Image A nalysis and Automated Cartography. *Communication. Assococi ation for Computing Machinery(ACM)*, 24(6): 381-395.

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

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

Hartley, R., and Zisserman, A., 2000. *Multiple View Geometry i n Computer Vision*. Cambridge University Press.

Hong, S.-M., 2002. Steady-state Analysis of Computational Loa d in Correlation-based Image Tracking. IEE Proceedings –Visio n, Image and Signal Processing, Vol. 149, No. 3.

Lim, S., and El Gamal, A., 2001. Optical Flow Estimation Usin g High Frame Rate Sequence, *Proceedings 2001 International Conference on Image Processing*, Thessaloniki, Greece, Oct. 7-10, Vol. 2, pp. 925-928.

Nguyen, H.T., Worring, M., and Boomgaard, M., 2001. Occlusi on Robust Adaptive Template Tracking. *8th IEEE International Conference on Computer Vision*, Vancouver, Canada, July 9-12, Vol. 1, pp. 678-683.

Pritchett, P., and Zisserman, A., 1998. Wide Baseline Stereo M atching, *6th International Conference on Computer Vision*, Bo mbay, India, pp. 754-760.

Tekalp, M., 1995. *Digital Video Processing*. Prentice Hall.

Trucco, E., and Verri, A., 1998. *Introductory Techniques for 3-D Computer Vision.* Prentice Hall.

Van Gool, L., Tuytelaars, T., Ferrari, V., Strecha, C., Vanden W yngaerd, J., Vergauwen, M., 2002. 3D Modeling and Registrati on Under Wide Baseline Conditions, *IASPRS Commission III*, p p. A3-A14.

ACKNOWNLEDGEMENTS

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.