Robust estimation of target scale by removing outlier motion vectors using MAD

J.W. Park and Y.S. Moon✉

A robust estimation method for target scale using median absolute deviation (MAD) for reducing outlier motion vectors is proposed. To remove outlier motion vectors, the proposed method first uses normalised cross-correlation (NCC) and forward–backward (FB) filters. Then, the MAD filter removes the remaining outlier motion vectors. After filtering, inlier motion vectors are used to estimate a scale change ratio of the target. The estimated scale change ratio is evaluated by the MAD of scale change ratios to guarantee reliability of estimation.

Introduction: In object tracking, scale estimation of a target is one of the major issues. Usually, the tracker based on the optical flow estimates the scale change ratio of the target using motion vectors calculated by optical flow. Although optical flow calculates motion vectors, errors of optical flow may increase because of blur, illumination change, camera motion and target motion. Therefore some motion vectors contain serious errors and are classed as outliers. These outliers disturb an accurate scale estimation of the target and it causes a gradual drift in tracking. To overcome this problem, Kalal *et al.* proposed the median flow tracker [1], which is the main tracker of the tracking-learning-detection (TLD) [2], which is one of the state-of-the-art trackers. Median flow tracker filters out outliers by normalised cross-correlation (NCC) and forward–backward (FB) filters effectively. Sometimes, however, rapid target motion or sudden illumination change significantly increases optical flow calculation errors. In this case, the median flow tracker fails to filter out outliers sufficiently and the remaining outliers cause an incorrect estimation of target scale, resulting in a gradual drift.

In this Letter, we propose a robust estimation of target scale using the median absolute deviation (MAD) [3]. The proposed method is composed of an outlier removal and evaluation of estimated scale change ratios. Outlier removal is performed using NCC and FB filtering, followed by MAD filtering. The proposed method first removes outlier motion vectors using two filtering phases and estimates the scale change ratio of the target. Finally, the reliability of the estimated scale change ratio in the bounding box is estimated. Fig. 1 shows the flowchart of the proposed method.

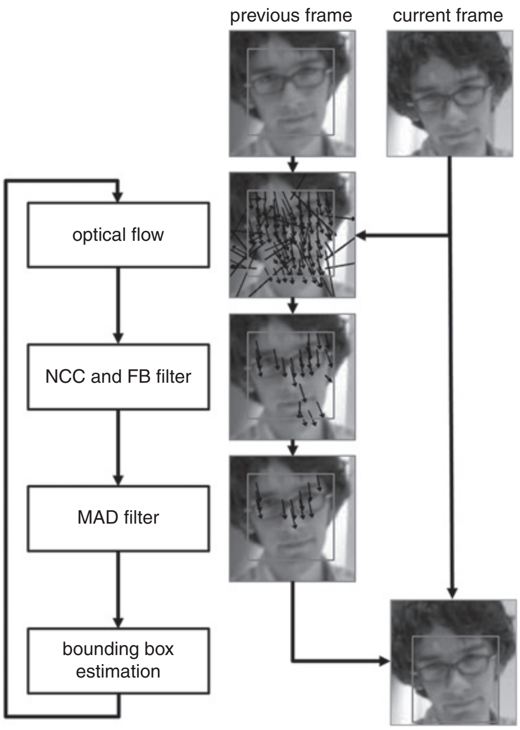

Fig. 1 *Flowchart of proposed method*

Novelty of proposed method: The framework of the proposed method is Park's method [4]. Reference [4] measures the motion vector error and the scale change ratio error to estimate the scale of the target. Reference [4] imitates the Kalal method, median flow tracker. However, the proposed method is based on MAD which has a mathematical basis [5]. MAD has been studied for a long time and verified by many researches. The logical explication of the proposed method is not only simpler than [4], but also guarantees the reliable estimation stochastically because of the proven study of MAD.

Outlier removal: All motion vectors are calculated by optical flow. An outlier motion vector means a motion vector that has a different direction or magnitude against the target motion. Outlier removal is performed by NCC and FB filtering and MAD filtering. NCC and FB filtering is the same as the Kalal method [1]. It filters out most outliers. However, some outliers still remain after the filtering.

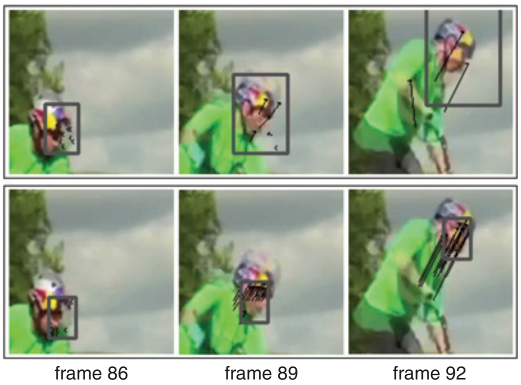

Fig. 2 *Cropped images of tracking results in Biker sequence*
Top row: Kalal's method; bottom row: proposed method

MAD filtering is performed after NCC and FB filtering to remove the remaining outliers. MAD is a robust measure of the variability of univariate samples of quantitative data. MAD is calculated as

$$M(x_i) = m(|x_i - m(x_i)|), \quad 1 \le i \le N \tag{1}$$

where $M(\cdot)$ is the MAD function; x_i is an element of univariate data; $m(\cdot)$ is a median function and N is the cardinal number of x_i. MAD is defined as the median of the absolute deviations from the median of a set. The MAD filter classifies inlier motion vectors from motion vectors filtered by NCC and FB filters. The MAD filter is designed as

$$f_{\rm M}(v_i) = \frac{|\boldsymbol{v}_i - m(\boldsymbol{v}_i)|}{M(\boldsymbol{v}_i)}, \quad 1 \le i \le N \tag{2}$$

where $f_{\rm M}(\cdot)$ is the MAD filter and $\boldsymbol{v}_i$ is the motion vector filtered by NCC and FB filters. If $f_{\rm M}(\boldsymbol{v}_i)$ is ≤ 1, the proposed method decides that $\boldsymbol{v}_i$ is an inlier motion vector; otherwise, it is an outlier motion vector. In Fig. 1, most of the motion vectors filtered by NCC and FB filters have consistency in terms of direction and magnitude; however, some motion vectors have dissimilar direction and magnitude. The MAD filter classifies dissimilar motion vectors as outliers and removes them.

Evaluation of estimated scale change ratio: Scale change ratio is calculated as

$$\begin{gathered} R_{ij} = \frac{D(p_i^{\rm e}, p_j^{\rm e})}{D(p_i^{\rm s}, p_j^{\rm s})}, \quad i \ne j, \quad 1 \le i, \ j \le K \\ (p_i^{\rm s}, p_j^{\rm e}) \in \boldsymbol{v}_i, \quad \text{for } f_{\rm M}(\boldsymbol{v}_i) \le 1 \end{gathered} \tag{3}$$

where $\boldsymbol{v}_i$ is the inlier motion vector of the MAD filter; $p_i^{\rm s}$ is a start point of $\boldsymbol{v}_i$; $p_i^{\rm e}$ is an endpoint of $\boldsymbol{v}_i$ and R_{ij} is the scale change ratio between $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$. $D(\cdot)$ is a distance function and (x_i, y_i) is the coordinate of p_i as

$$D(p_i, p_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, \quad (x_i, y_i) \in p_i \tag{4}$$

The proposed method calculates the median of scale change ratios and the median is used as an estimated scale change ratio of the target. The estimated scale change ratio is evaluated by the MAD of scale change ratio as in (5) because the MAD filter may fail to remove enough outliers when heavy outliers occur abruptly:

$$G(R_{ij}) = \frac{M(R_{ij})}{T}, \quad i \neq j, \quad 1 \leq i, \quad j \leq K, \qquad 0 < T < 1 \qquad (5)$$

where $G(\cdot)$ is an evaluation function of the estimated scale change ratio and T is a threshold. T is a real number and used to measure the reliability of the estimated scale change ratio. We fix T at 0.1 in the whole experiment. If $G(R_{ij})$ is <1, the estimated scale change ratio is used to calculate target scale at the next frame. Otherwise, the proposed method keeps the scale of target at the current frame to the next frame.

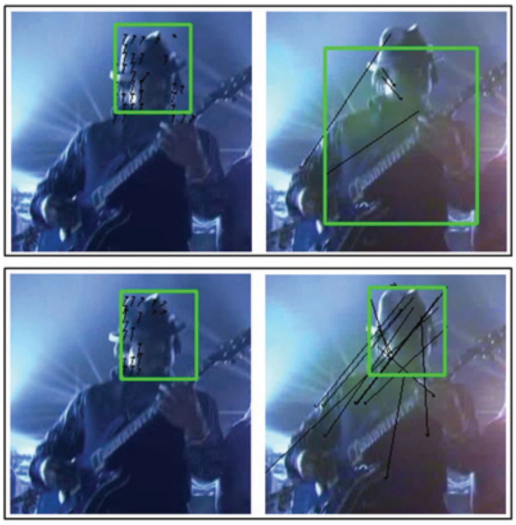

Fig. 3 *Cropped images of tracking results in Shaking sequence*
Top row: Kalal's method; bottom row: proposed method

Experimental results: We compare the performance of the proposed method with Kalal's method [1], which is a tracker based on the motion vector, similar to the proposed method. Kalal's method is known to be the state-of-the-art tracker TLD [2]. We use video sequences with rapid motions and dynamic illumination change because rapid motions and dynamic illumination change cause a considerable increase of incorrect motion vectors.

Table 1: Comparative results of tracking performance

Sequence name	Number of frames	Kalal's method (frame number)	Proposed method (frame number)
Animal	71	11	71
Biker	180	91	180
Bolt	293	54	125
Jumping	313	96	313
Kitesurf	84	84	84
Shaking	365	58	365

Fig. 2 shows the frames when the biker is jumping. The biker head moves up rapidly and the bounding box indicates the position and scale of the biker head at those frames. At frame 86 in Fig. 2, it is shown that the sizes of the bounding boxes for the two methods are almost the same. Kalal's method (top row of Fig. 2) estimates the centre of the target correctly. However, the scale of the target is estimated to be much larger than the correct scale because of incorrect motion vectors caused by the rapid motion. On the other hand, the proposed method (bottom row of Fig. 2) estimates the size of the target stably because the MAD filter filters out incorrect motion vectors and the scale is estimated using only inlier vectors.

Fig. 3 shows the result of scale estimation when the illumination changes. The target is the head of a singer. At frame 58, both the methods have heavy outlier motion vectors. As a result, Kalal's method estimates an incorrect scale of the target. On the other hand, the proposed method keeps the scale of the target because the result of the evaluation function is more than 1. It means that the estimated scale change ratio at frame 58 is significantly changed over the threshold; therefore the estimated result is not reliable.

Table 1 shows the last frame that each method succeeded to track the target correctly for six popular sequences (jumping [2], biker, bolt, kitesurf [6], animal and shaking [7]). The bounding box at the first frame is initialised manually. If the target area covers more than 50% of the bounding box, we decide a tracking success. The proposed method tracks the target correctly for longer than the conventional method.

Conclusion: In this Letter, we have proposed a robust estimation method for target scale using MAD. The proposed method filters out motion vectors that have a high distance error using the MAD filter. For an accurate estimation of the scale change ratio, the proposed method calculates the MAD of scale change ratio errors and uses this MAD to guarantee the accuracy of the scale change ratio for the next frame. Experimental results show that the proposed method achieves better performance of target tracking than the conventional method. The first contribution of the proposed method is a robust filtering when heavy inconsistent motion vectors occur and the second is a stable estimation of the scale change of a target in video sequences.

28 October 2014
doi: 10.1049/el.2014.3821
One or more of the Figures in this Letter are available in colour online.

J.W. Park and Y.S. Moon (*Department of Computer Science and Engineering, Computer Vision and Pattern Recognition Laboratory, Hanyang University, Republic of Korea*)
✉ E-mail: ysmoon@hanyang.ac.kr

References

1 Kalal, Z., Mikolajczyk, K., and Matas, J.: 'Forward–backward error: automatic detection of tracking failures'. IEEE Int. Conf. on Pattern Recognition, Istanbul, Turkey, August 2010, pp. 2756–2759
2 Kalal, Z., Mikolajczyk, K., and Matas, J.: 'Tracking-learning-detection', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011, **34**, (7), pp. 1409–1422
3 Hampel, F.R.: 'The influence curve and its role in robust estimation', *J. Am. Stat. Assoc.*, 1974, **69**, pp. 383–393
4 Park, J.W., and Moon, Y.S.: 'Robust tracking with accurate scale estimation using errors of motion vectors'. IEEE Int. Symp. Consumer Electronics, Jelu Island, South Korea, June 2014, pp. 1–2
5 Daszykowskia, M., Kaczmarekb, K., Vander Heydenc, Y., and Walczak, B.: 'Robust statistics in data analysis – a review: basic concepts', *Chemometr. Intell. Lab. Syst.*, 2007, **85**, (2), pp. 203–219
6 Zhang, K., Zhang, L., and Yang, M.: 'Real-time compressive tracking'. Proc. of European Conf. on Computer Vision, Florence, Italy, October 2012, pp. 864–877
7 Kwon, J., and Lee, K.: 'Visual tracking decomposition'. Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, San Francisco, CA, USA, June 2010, pp. 1269–1276