Tracking of Moving Objects With Regeneration of Object Feature Points

Igor I. Lychkov
Information Systems and Telecommunications Department Bauman Moscow State Technical University
Moscow, Russian Federation
lychkovi@bmstu.ru

Alexander N. Alfimtsev
Information Systems and Telecommunications Department Bauman Moscow State Technical University
Moscow, Russian Federation
alfim@mail.ru

Sergey A. Sakulin
Information Systems and Telecommunications Department Bauman Moscow State Technical University
Moscow, Russian Federation
ss141291@yandex.ru

***Abstract*—This paper concerns moving object tracking in the videos, based on sparse optical flow technique. Current optical flow tracking methods suffer from feature points loss. We extended an existing sparse optical flow tracking method with a new function for automatic feature points' recovery that uses biological regeneration principle. Besides, we improved the tracking method to deal with object rotation and scaling transformations. We applied the improved tracking method to a real video and noticed acceptable tracking performance. Our experiment showed that the proposed tracking method with feature points' recovery provides higher tracking accuracy than the original tracking method without feature points recovery when the moving object is partially occluded by an obstacle.**

Keywords—moving object tracking, optical flow, object occlusion, feature points detection, biological regeneration

I. Introduction

Automatic tracking of moving objects in videos is an important task that arises in such fields as security, defense, robotics and etc. It consists in computation of moving object position in a sequence of video frames given its position in the first video frame of the sequence. Automatic tracking of moving objects is a challenging task for videos with object occlusion and poor illumination.

II. Literature Review

One can distinguish five main approaches in moving object tracking: blob tracking, contour-based tracking, optical flow tracking, model-based tracking and neural network tracking. Blob tracking [1–6] uses background subtraction techniques and is not appropriate for videos recorded with a moving camera. Neural network tracking [7–10] requires large training datasets that can be found only for certain object categories and only for conventional videos. It is difficult to prepare large training datasets for videos produces by infra-red cameras or electronic optical converters of night vision devices. Contour-based tracking [11, 12], 3D model-based tracking [13, 14] and dense optical flow tracking [15–17] require large computational efforts therefore these methods are not suitable for real-time applications. In this work we improve an existing sparse optical flow tracking method [18] based on Lucas-Kanade optical flow [19] that eliminates training and is suitable for real-time applications [20].

III. Problem Statement

Sparse optical flow tracking [21–23] consists in detection of salient feature points in a video frame and subsequent tracking of the detected feature points in series of video frames. Due to imperfections of tracking algorithms some feature points can be lost. Initially, several feature points can be detected within the contour of a moving object. A considerable part of image patch around such a feature point belongs to background. While being tracked this feature point can leave the object region and enter the background region thus introducing uncertainty in moving object tracking. Besides, non-transparent obstacles can hide some parts of a moving object and make it difficult to track the moving object in a proper way. In order to increase tracking accuracy we propose a method for automatic recovery of the lost feature points by means of detection of new feature points within the object region. The main idea of the recovery method is based on exploitation of biological regeneration principles.

IV. Feature Points Recovery Based on Biological Regeneration

Regeneration is a special ability of biological organism to recover some lost parts of its body. Consider three known principles of biological regeneration [24]: epimorphosis, moralphallaxis and endomorphosis. Epimorphosis consists in reduction of body tissues adjacent to a wounded surface, growth of cells and their subsequent differentiation. Moralphallaxis is mostly similar to epimorphosis except that the rest cells of organism are to be rearranged. Endomorphosis consists in recovery of damaged organ mass based on division and hypertrophy of cells.

The aforementioned regeneration principles can be adapted to lost feature points recovery in object tracking task as follows. Epimorphosis keeps the original geometric parameters of moving objects. The major criterion for assignment of a feature point to a moving object is the distance from this feature point to the nearest object center. The main idea of moralphallaxis is mostly similar to epimorphosis except that a further optimization is implemented in order to reduce the number of the poorly tracked feature points. Endomorphosis considers the number of the feature points that are currently being tracked within the object bounding box.

This work was supported by RF Ministry of Education and Science R & D State project №2.5048.2017/8.9.

978-1-5386-7386-7/18/$31.00 ©2018 IEEE

In this work we propose to use epimorphosis as a basic principle for the feature points recovery. User specifies threshold for a number of the feature points to be tracked as an input parameter of the method. The image feature points are tracked in every video frame using Lucas-Kanade sparse optical flow [19] and the objects' bounding boxes are computed by means of a moving object tracking method presented in the following section. When the number of the feature points within the object bounding box decreases below the user-specified threshold new feature points will be detected in the following way. At first, we apply Shi-Tomasi feature point detector [25] to detect new feature points within the object bounding box. At second, we reject the feature points that are located at distances less than 5 pixels to the original object feature points currently being tracked. Finally, we select the necessary number of the feature points which are the closest to the center of the object bounding box in current video frame.

V. Method for Moving Object Tracking in a Video

The critical problem of feature point tracking based on sparse optical flow consists in migration of distinct feature points from moving object image region to background image region that introduces errors in tracking results. In order to achieve reliable tracking of moving objects we need to distinguish and reject the lost feature points.

Method for moving object tracking based on relative motion network [18] makes it possible to match the observed object positions in current video frame to the object trajectories in a sequence of video frames according to their relative displacements. Maximum likelihood principle is used to compose the optimal matching.

The original tracking method [18] assumes that distinct object is represented with a single feature point in the video frame, deals only with translational motion and does not consider rotation and scaling of object image within the frame during its motion. In this work we propose a modification for the original tracking method to deal with object rotation and scaling.

The input data for the method include the set of observations $Z_t = \{z_t^k\}$ computed using optical flow technique, where t is a video frame index, k is a feature point index, $z_t^k = [z_{ut}^k, z_{vt}^k]$ is an observation vector with co-ordinates z_{ut}^k and z_{vt}^k.

Current state of moving objects to be tracked is represented with a set of states $X_t = \{(x_t^i, h_t^i)\}$, where t is a video frame index, i is a moving object index ($i = 0$ stands for background), $x_t^i = [x_{ut}^i, x_{vt}^i]$ is a vector of the i-th object center within the t-th video frame, $h_t^i = [h_{ut}^i, h_{vt}^i]$ represents a size of the i-th object bounding box.

In order to compute the object center position and relative displacements of distinct feature points we compose a relative motion model $R_t = \{(m_j^i, n_j^i), \pi_i\}$, where (m_j^i, n_j^i), $j = 1 : 2N$ is a list of indices of reference feature point pairs for the i-th object, π_i stands for indices of the main reference feature point pair $(m_{\pi i}^i, n_{\pi i}^i)$ that is used to track the i-th object on the t-th video frame. For the initial video frame $t = 0$ the lists of indices are composed using random sampling of $2N$ feature point pairs from the total set of the detected feature points.

Then we match the feature points to the moving objects and find the main reference feature points of objects using the maximum likelihood principle:

$$p(Z_t \mid X_t, R_{t-1}) = \max_{A,\pi} \prod_{i,k} p(z_t^k \mid x_t^i, R_{t-1}^{\pi})^{a^{i,k}}, \tag{1}$$

where R_{t-1} is the relative motion model for the previous video frame, R_{t-1}^{π} is the relative motion model with modified indices of the main reference feature point pairs of all objects, $A = \{a^{i,k}\}$ and $a^{i,k} \in \{0, 1\}$ represent the matching correspondences of the feature points to the objects, $p(z_t^k \mid x_t^i, R_{t-1}^{\pi})$ is a partial likelihood function.

In order to simplify the optimization problem one can apply the logarithm function to the right side of (1):

$$p(Z_t \mid X_t, R_{t-1}) \sim \max_{A,\pi} \prod_{i,k} a^{i,k} \ln p(z_t^k \mid x_t^i, R_{t-1}^{\pi}). \tag{2}$$

The relative motion model is adjusted in the following way. The solution of optimization problem (2) gives the optimal values of the matrix A elements. One can select the best reference feature point pair of the model R_{t-1} under condition of global optimum (2) and $(N - 1)$ best reference feature points under condition of local maximum likelihood for the i-th object:

$$\pi_{i\max} = \arg\max_{A,\pi i} \prod_{i,k} a^{i,k} \ln p(z_t^k \mid x_t^i, R_{t-1}^{\pi}). \tag{3}$$

The rest N pairs are selected using random sampling from a set of the feature points of the i-th object in the t-th video frame.

The partial likelihood function is assigned to:

$$p(z_t^k \mid x_t^i, R_{t-1}^{\pi}) = G(z_\tau^k, y_\tau^k), \tag{4}$$

where G is a two-dimensional Gaussian

$$G(v, v_0) = \exp\left\{-\frac{1}{2}(v - v_0)^{\mathrm{T}} C (v - v_0)\right\} \tag{5}$$

with a covariance matrix

$$C = \begin{bmatrix} \sigma_u^2 & \sigma_u\sigma_v \\ \sigma_u\sigma_v & \sigma_v^2 \end{bmatrix} \tag{6}$$

and standard deviations for distinct co-ordinates:

$$3\sigma_u = \frac{1}{2} h^i_{u\tau}, \quad 3\sigma_v = \frac{1}{2} h^i_{v\tau}, \tag{7}$$

which depend on the i-th object size at time τ of the k-th feature point detection.

$y_\tau^k = [y_{u\tau}^k, y_{v\tau}^k]^T$ in (4) is a predicted position of the k-th feature point in the τ-th video frame assuming that the k-th feature point is assigned to the i-th object and the relative motion model of the i-th object feature points is affine transform model.

The affine transform matrix $M_{ti} \in R^{2\times3}$ for the i-th moving object is determined using the following system of equations:

$$\begin{cases} [z^{p1}_{u\tau} \quad z^{p1}_{v\tau}]^T = M_{ti} \cdot [z^{p1}_{ut} \quad z^{p1}_{vt} \quad 1]^T \\ [z^{p2}_{u\tau} \quad z^{p2}_{v\tau}]^T = M_{ti} \cdot [z^{p2}_{ut} \quad z^{p2}_{vt} \quad 1]^T , \\ [z^{p3}_{u\tau} \quad z^{p3}_{v\tau}]^T = M_{ti} \cdot [z^{p3}_{ut} \quad z^{p3}_{vt} \quad 1]^T \end{cases} \tag{8}$$

where $p1$, $p2$, $p3$ are indices of the feature points assigned to the i-th moving object.

The predicted position of the k-th feature point in the τ-th video frame is computed using the affine transform matrix M_{ti}:

$$y^k_\tau = M_{ti} \cdot [z^k_{ut} \quad z^k_{vt} \quad 1]^T. \tag{9}$$

A feature point that moved from the moving object region to a background region will go far away from the moving object. Its partial likelihood function (4) takes greater values if the feature point is assigned to the background and smaller values if the feature point is assigned to the moving object. Thus the optimal solution for (2) gives us a reasonable matching of the feature points to the moving objects or to the background.

VI. Experimental Setup and Results

We have recorded an original video with a frame rate of 25 fps, duration of 50 seconds and resolution 720×576 pixels that represents road traffic on a highway. We used the proposed object tracking method to extract reference trajectories for 14 moving cars from the original video. The initial trajectory points were specified manually. In order to simulate occlusion of the moving objects with an obstacle we put 4 vertical gray stripes (each stripe is 12 pixels width) on all video frames thus produced a modified video.

We applied the proposed object tracking method to the modified video thus produced resulting trajectories for 14 moving cars in the modified video for two cases. In the first case the feature points recovery was not implemented and in the second case the feature points recovery was implemented according to the proposed method of regeneration. Initial trajectory point co-ordinates were assigned to the same values as for the corresponding reference trajectories. Then we matched the resulting trajectories to the reference trajectories and calculated two performance metrics: a percentage of successfully tracked objects and a tracking deviation. A moving object was considered to be successfully tracked if the distance of its position in the resulting trajectory from its position in the reference trajectory does not exceed the specified maximum distance threshold for each video frame. The maximum distance threshold was specified manually in percents of the object bounding box size. The percentage of successfully tracked objects was computed as the number of successfully tracked objects divided by the total number of reference trajectories. The tracking deviation was computed as an average standard deviation of the resulting trajectories from the reference trajectories among all successfully tracked objects. The number of the object feature points to be tracked was manually set to 10. The experimental results are presented in Table I.

TABLE I. EXPERIMENTAL RESULTS

Tracking Method	**Maximum Distance Threshold is set to 50% of Object Bounding Box Size**		**Maximum Distance Threshold is set to 75% of Object Bounding Box Size**	
	Percentage of successfully tracked objects	***Tracking deviation (pixels)***	***Percentage of successfully tracked objects***	***Tracking deviation (pixels)***
Without lost feature points recovery	43 %	10,433	64 %	15,103
With lost feature points recovery	43 %	8,765	71 %	17,212

When the maximum distance threshold is set to 0,5 of the object bounding box size, tracking with lost feature points recovery provides lower tracking deviation than tracking without lost feature points recovery. When the maximum distance threshold is set to 0,75 of the object bounding box size, tracking with lost feature points recovery provides higher percentage of successfully tracked objects than tracking without lost feature points recovery. Examples of the processed video frames with bounding boxes computed for a distinct moving object are presented in Fig. 1, Fig. 2, and Fig. 3.

VII. Conclusions

In this work we extended an existing method for moving objects tracking in videos with a function for automatic lost feature points recovery based on biological regeneration principle. Experiment carried out on a real video showed that automatic lost feature points recovery makes it possible to increase the tracking accuracy in case of a moving object occluded by a stationary obstacle.

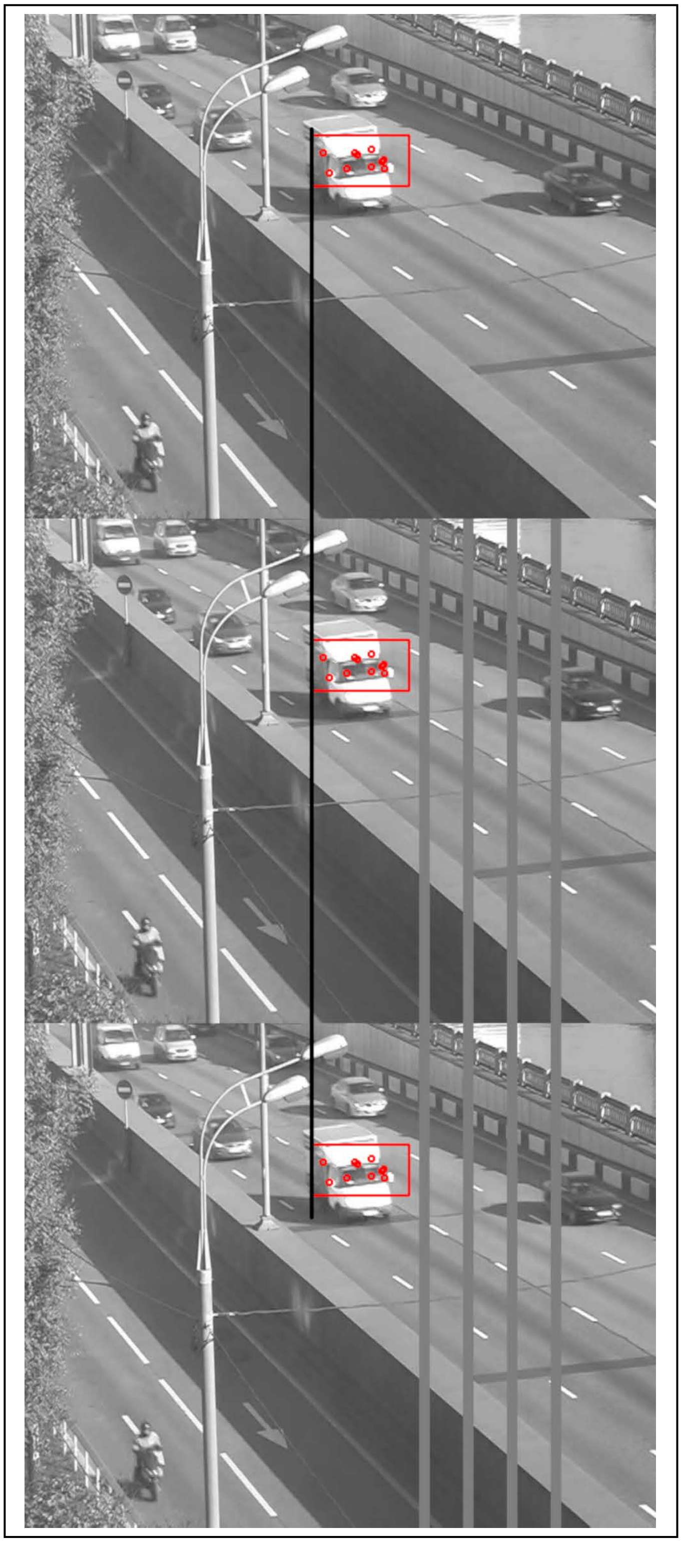

Fig. 1. Object bounding boxes in the 985-th video frame: reference trajectory (upper image); resulting trajectories computed without (middle image) and with (lower image) feature points recovery.

Fig. 2. Object bounding boxes in the 1005-th video frame: reference trajectory (upper image); resulting trajectories computed without (middle image) and with (lower image) feature points recovery.

Fig. 3. Object bounding boxes in the 1021-st video frame: reference trajectory (upper image); resulting trajectories computed without (middle image) and with (lower image) feature points recovery.

The use of affine motion model makes it possible to deal with rotation and scaling transformations of a moving object while tracking.

ACKNOWLEDGMENT

This work was supported by RF Ministry of Education and Science R & D State project №2.5048.2017/8.9.

REFERENCES

[1] J. Wu, "Study of multiple moving targets' detection in fisheye video based on the moving blob model," Multimedia Tools and Applications, pp. 1–20, 2018.

[2] L. Yao and M. Ling, "An improved mixture-of-Gaussians background model with frame difference and blob tracking in video stream," The Scientific World Journal, pp. 1–9, 2014.

[3] Y. Zhao, B. Zhang, and X. Zhang, "Meanshift blob tracking with target model adaptive update," Proceedings of the 33rd Chinese Control Conference, pp. 4831–4835, 2014.

[4] A. R. Taranyan, V. V. Devyatkov, and A. N. Alfimtsev, "Selective covariance-based human localization, classification and tracking in video streams from multiple cameras," Proceedings of the 9th International Conference on Bioinformatics Models, Methods and Algorithms, Part of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies, vol. 3, pp. 81–88, 2018.

[5] A. A. Morozov, A. F. Polupanov, V. E. Antciperov, A. Vaish, I. I. Lychkov, A. N. Alfimtsev, and V. V. Deviatkov, "Development of Concurrent Object-oriented Logic Programming System to Intelligent Monitoring of Anomalous Human Activities," Proceedings of the 7th International Conference on Biomedical Electronics and Devices, Part of the 7th International Joint Conference on Biomedical Engineering Systems and Technologies, pp. 53–62, 2014.

[6] V. V. Deviatkov and I. I. Lychkov, "Recognition of Dynamical Situations on the Basis of Fuzzy Finite State Machines," Proceedings of IADIS International Conferences Computer Graphics, Visualization, Computer Vision and Image Processing 2017 and Big Data Analytics, Data Mining and Computational Intelligence 2017, pp. 103–109, 2017.

[7] E. Price, G. Lawless, H. H. Bülthoff, M. Black, and A. Ahmad, "Deep Neural Network-based Cooperative Visual Tracking through Multiple Micro Aerial Vehicles," arXiv preprint arXiv:1802.01346, 2018.

[8] M. Danelljan, A. Robinson, F. S. Khan, and M. Felsberg, "Beyond correlation filters: Learning continuous convolution operators for visual tracking," Proceedings of European Conference on Computer Vision, pp. 472–488, 2016.

[9] G. Ning, "Spatially supervised recurrent convolutional neural networks for visual object tracking," Proceedings of IEEE International Symposium on Circuits and Systems, pp. 1–4, 2017.

[10] Yu. S. Fedorenko and Yu. E. Gapanyuk, "Multilevel neural net adaptive models using the metagraph approach," Optical Memory and Neural Networks (Information Optics), vol. 25, pp. 228–235, 2016.

[11] M. S. Allili and D. Ziou, "Object contour tracking in videos by using adaptive mixture models and shape priors," Proceedings of the International Symposium on Computational Modelling of Objects Represented in Images, p. 47, 2006.

[12] P. Christiansen, K. A. Steen, R. N. Jørgensen, and H. Karstoft, "Automated detection and recognition of wildlife using thermal cameras," Sensors, vol. 14, pp. 13778–13793, 2014.

[13] C. Y. Ren, V. A. Prisacariu, O. Kähler, I. D. Reid, and D. W. Murray, "Real-time tracking of single and multiple objects from depth-colour imagery using 3D signed distance functions," International Journal of Computer Vision, vol. 124, pp. 80–95, 2017.

[14] K. Pauwels and D. Kragic, "Simtrack: A simulation-based framework for scalable real-time object pose detection and tracking," Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1300–1307, 2015.

[15] J. Janai, F. Güney, J. Wulff, M. Black, and A. Geiger, "Slow flow: Exploiting high-speed cameras for accurate and diverse optical flow reference data," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 3597–3607, 2017.

[16] P. Ochs, J. Malik, and T. Brox, “Segmentation of moving objects by long term video analysis,” IEEE transactions on pattern analysis and machine intelligence, vol. 36, pp. 1187–1200, 2014.

[17] Y. Hua, K. Alahari, and C. Schmid, “Occlusion and motion reasoning for long-term tracking,” Proceedings of European Conference on Computer Vision, pp. 172–187, 2014.

[18] J. H. Yoon, M. H. Yang, J. Lim, and K. J. Yoon, “Bayesian multi-object tracking using motion context from multiple objects,” Proceedings of IEEE Winter Conference on Applications of Computer Vision, pp. 33–40, 2015.

[19] J. Y. Bouguet, “Pyramidal implementation of the affine Lucas-Kanade feature tracker description of the algorithm,” Technical Report, Intel Corporation, 2001.

[20] N. A. Yakusheva, A. V. Proletarsky, and M. Basarab, “Timing Analysis for LOT-Based Vehicle-Pedestrian Collision Avoidance for NLOS Conditions,” Proceedings of the 3rd International Conference on Smart and Sustainable Technologies, art. no. 8448355, pp. 1–5, 2018.

[21] L. Dan, “Moving object tracking method based on improved Lucas-Kanade sparse optical flow algorithm,” Proceedings of IEEE International Smart Cities Conference, pp. 1–5, 2017.

[22] D. M. Woo and A. N. Hong, “An integrated solution to motion tracking of moving vehicles using optical flow,” Int. Journal of Computing, Communication and Instrumentation Eng., vol. 1, pp. 60–63, 2014.

[23] J. P. Jodoin, G. A. Bilodeau, and N. Saunier, “Tracking all road users at multimodal urban traffic intersections,” IEEE Transactions on Intelligent Transportation Systems, vol. 17, pp. 3241–3251, 2016.

[24] K. Agata, Y. Saito, and E. Nakajima, “Unifying principles of regeneration I: Epimorphosis versus morphallaxis,” Development, growth & differentiation, vol. 49, pp. 73–78, 2007.

[25] J. Shi, “Good features to track,” Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994, pp. 593–600, 1994.