DYNAMIC KEYPOINT-BASED ALGORITHM OF OBJECT TRACKING

A. I. Morgacheva [a, b], V. A. Kulikov [b], V. P. Kosykh [b]

[a] Novosibirsk State University, Novosibirsk - a.morgacheva@g.nsu.ru
[b] Institute of Automation and Electrometry SB RAS, Novosibirsk - kulikov.victor@gmail.com, kosych@iae.nsk.su

Commission II, WG II/10

KEY WORDS: Tracking , Keypoints, Dynamic object model, FRiS,

ABSTRACT:

The model of the observed object plays the key role in the task of object tracking. Models as a set of image parts, in particular, keypoints, is more resistant to the changes in shape, texture, angle of view, because local changes apply only to specific parts of the object. On the other hand, any model requires updating as the appearance of the object changes with respect to the camera. In this paper, we propose a dynamic (time-varying) model, based on a set of keypoints. To update the data this model uses the algorithm of rating keypoints and the decision rule, based on a Function of Rival Similarity (FRiS). As a result, at the test set of image sequences the improvement was achieved on average by 9.3% compared to the original algorithm. On some sequences, the improvement was 16% compared to the original algorithm.

1. INTRODUCTION

Object tracking is in great demand nowadays and used in various applications, including surveillance. Tracking an object on a set of images from different cameras is quite difficult task, therefore, tracking programs are used to facilitate the work of the operator.

The task of tracking for a priori unknown objects (object tracking) has a considerable interest in the field of computer vision. Under a priori unknown object, it is meant that the input trace algorithm has as an input only the area containing the object in the first frame of sequence. The object tracking problem is to determine the position of the object in subsequent frames, knowing its location on the first frame.

During objects tracking, the algorithm encounters a number of difficulties: low contrast between the object and the background, overlapping the object of interest with other objects, temporary disappearance of the object from the frame, changes in lighting, shape, viewing angle or texture of the object.

The key role in the tracking task is the representation (model) of the object in the algorithm. Object models as a set of parts are more resistant because local changes apply only to certain parts of the object.

The CMT (Nebehay, G., 2014) algorithm suggests to use a model of an object as a set of keypoints. When the form and viewing angle of object are changed, the set of keypoints which is describing the object or tracking also changes. But the changes are not reflected in the model. Therefore, it is proposed to modify the algorithm CMT using a model with dynamically modified set of keypoints.

2. CMT ALGORITHM

CMT algorithm was chosen as the basis of our work. This algorithm uses the representation of the object in the form of a set of BRISK (Leutenegger, S., 2011) keypoints. CMT was chosen for the combination of accuracy, speed and compact representation of the model.

The model in the algorithm consists of the center of mass, the coordinates of the keypoints relative to the center of mass, the data of their relative location and descriptors. This algorithm uses the combined operation of a tracker and a detector. The tracker determines the position of object points in the new frame, using the computation of forward and backward optical flow (Lucas, B. D., 1981)] that allows more accurately determine the new position of the object points. Based on the new location of the points, the displacement of the object, the change in its size and the angle of rotation are calculated. To calculate the center of the object and removing outliers applies clustering. Then the detector compares the descriptors of the frame with the descriptors in the object model and specifies the result of the tracker's work. Based on the obtained data on the model change, a new position of the object area on the frame is calculated.

3. DYNAMIC MODEL

The sequence of frames of the input video $I_1 \dots I_T$ and the area b_1, which is bounding the object in the first frame of the sequence are given. The task of the algorithm is to estimate the location of this area on each subsequent frame or to determine that the object is not visible on the current frame. In the CMT algorithm, parameters such as its center of mass μ, scale S and angle of rotation θ are used to estimate the location of the object.

The object of observation is represented by the center of mass μ, the set of object keypoints $\mu O = \{r_i\}_{i=1}^{N}$, where r is the coordinate of each object keypoints with respect to the center of mass. To estimate the change in scale and rotation, the coordinate difference vectors for each pair of points of the object are computed, in the model they are represented by a matrix of the lengths of these vectors $L = \{l_{ij}\}_{i,j=1}^{N}$ and a matrix of angles in polar coordinates$L = \{l_{ij}\}_{i,j=1}^{N}$ $A = \{a_{ij}\}_{i,j=1}^{N}$. For the speed of calculations we used binary descriptors BRISK of dimension 512.

3.1 Ratings

In the process of tracking an object, the set of singular points describing the object can change quite strongly - old points can cease to be present on the object, new ones may appear. In accordance with the change in the object, it is necessary to change its model. To determine which points need to be added and which need to be deleted, a point rating system that determines the degree of confidence in each point is used.

At initial initialization, each object point is assigned a certain weight $w_i \epsilon W$, where W is the set of weights of all points of the object. From each frame I_k, all keypoints are extracted, their descriptors are mapped to the object model descriptors. If the presence of an object point on the new sequence frame is established, it is considered that such a point is more successful for tracking, and its weight increases.I_k When the maximum weight w_{max} is reached, the weight of the point no longer changes, since it is considered that such a point describes the object well enough.w_{max} If the point could not be found on the new frame, then its weight decreases. Thus, it is possible to get rid of the "random" points caused by noise or inaccuracy in the selection of the area of the object, and points that cease to be visible. The choice of initial weight and the criterion for adding a new point to the model are determined using FRiS (Zagorujko N. G., 2013).

3.2 FRiS decision rule

The point rating system should take into account the degree of confidence in each of the points in the object. As a measure of the degree of confidence in a point, it is used a function that estimates the degree of similarity of a given point to the points of the background and the object in the absolute scale. This function of rival similarity (FRiS) has the following properties:

- Locality property: the measure of similarity does not depend on the nature of the distribution of all points, but on the features of the distribution of points in the neighborhood of a given point.
- Normalization property: the function takes a maximum value of 1 when the point coincides with some point of the object, minus 1 when it coincides with a certain background point. In other cases, the function takes values from minus 1 to 1.
- The property of antisymmetry: the values of the point's similarity to the background and the object are related by the property of antisymmetry, at equal distances to the nearest point of the background and the object, the point will be equally similar to the background and the object, and the function will take the value 0.
- The property of invariance: the values of the measure of similarity are preserved for affine transformations.

Thus, FRiS allows to determine the degree of similarity of the point with the nearest point of the background in competition with the nearest point of the object. This measure of similarity reproduces the similarity assessment mechanism used by human, and is invariant to the distribution of keypoints.

The initial weight of the points of the object is calculated as follows: for each point, the distance d_b between the descriptor of this point and the nearest background descriptor f_b and the distance d_o between the point descriptor and the nearest object descriptor f_o are calculated. Then, the value of the FRiS

$$\alpha_i = \frac{d_{b_i}-d_{o_i}}{d_{b_i}+d_{o_i}} \quad (1)$$

$d_b f_b d_o f_o \alpha_i$ $\alpha_i = \frac{d_{b_i}-d_{o_i}}{d_{b_i}+d_{o_i}}$

is calculated for each point. As a result, the weight

$$w_i = F(\alpha_i) \quad (2)$$

$w_i = F(\alpha_i)$where F is a linear function with the following properties:

$$F(\alpha_i) = \begin{cases} w_{max}, & \text{if } \alpha_i = 1 \\ \frac{w_{max}}{k}, & \text{if } \alpha_i = -1 \end{cases} \quad (3)$$

where w_{max} is the maximum possible weight of a point and k is some natural coefficient.

Figure 1. New points of the object. Green shows the points added to the model.

The process of adding points to the model also uses FRiS. First, for each point in the frame, it is determined whether it belongs to an area defined as an object. If this is the case, the value of the FRiS function is calculated (1).: $\alpha_i = \frac{d_{b_i}-d_{o_i}}{d_{b_i}+d_{o_i}}$. If $\alpha \epsilon [t_1, t_2]$, where t_1, t_2 are given threshold values, and the distance d between this point and the point f_o in space $d f_o L_2$ is greater than a certain threshold D, then this frame point is considered to be new and it should be added to the model with some initial weight w_s.

After deleting and adding points, we need to recalculate the object model. Based on the new set of points, the center of mass of the object and the location of the points relative to it are recomputed. The corresponding rows and columns in matrices A and L are replacing.

4. RESULTS

4.1 Evaluation Methodology

To compare the proposed algorithm with the dynamic model, and without it we used the same set of test video sequences, as in the original paper.

There are a number of measures for quantifying the effectiveness of the tracking algorithms. As a measure K_1, showing the accuracy of the location of the object, we take the widely used relation (Klein, D. A., Schulz, D., Frintrop, S., & Cremers, A. B., 2010):

$$K_1 = \frac{S_0 \cap S_1}{S_0 \cup S_1} \quad (4)$$

$K_1 = \frac{S_0 \cap S_1}{S_0 \cup S_1}$ where S_0 is the area of the previously known rectangle containing the object, $S_1 S_0$ is the area of the rectangle, Found by the algorithm (Figure 2).

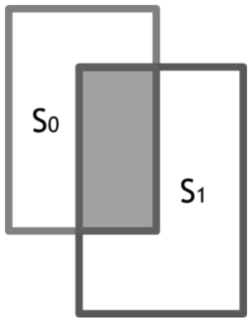

Figure 2. Comparison criterion K1.

Also considered such comparison criterion K_2, as the ratio of the number of frames where the object was able to detect T' to the total number of frames T:

$$K_2 = \frac{T'}{T} \tag{5}$$

To calculate this criterion, we take into account only those frames where the object is present in the field of view.

4.2 The influence of parameters on the quality of the tracking

We conducted experiments investigating the dependence of these criteria on various parameters of the algorithm. There were considered parameters: D - the threshold for the distance between the added point and the nearest point of the object and k - the coefficient of the initial weight of the point determining the function. The results of the experiments are shown in the graphs (Fig. 3-5).

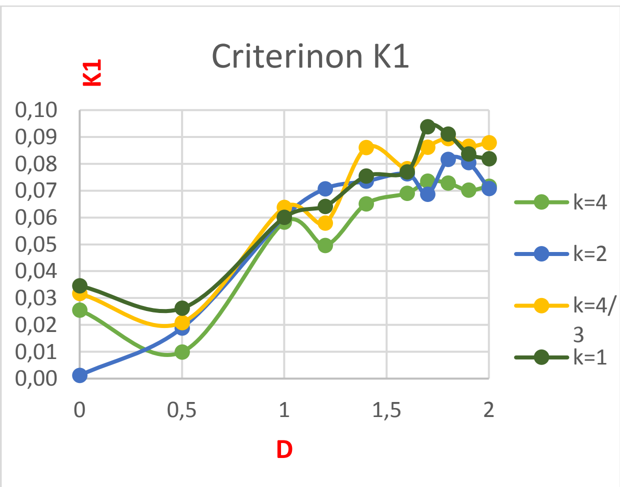

Figure 3a. Comparison criterion K_1. This figure shows an improvement in the algorithm compared to the original algorithm in fractions from 0 to 1. Four graphs illustrate the dependence of the criterion K_1 on D for different values of the parameter k giving us the initial weight of the points.

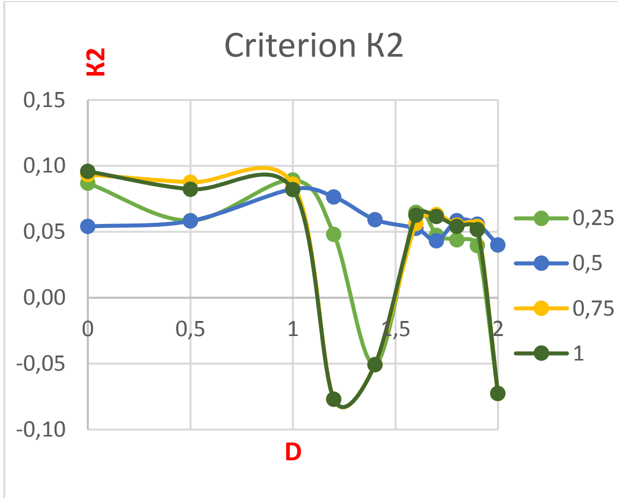

Figure 3b. Comparison criterion K_2. This figure shows an improvement in the algorithm compared to the original algorithm in fractions from -1 to 1. Four graphs illustrate the dependence of the criterion K_2 on D for different values of the parameter k giving us the initial weight of the points.

From the data of the experiments, it is evident that, according to the first criterion, the most optimal values are d = 1.7, k = 1. With these parameters, the average improvement was 9.37% for the first criterion, and for the second criterion 6.16%. At these values, the following results were obtained for different sequences (Fig. 4):

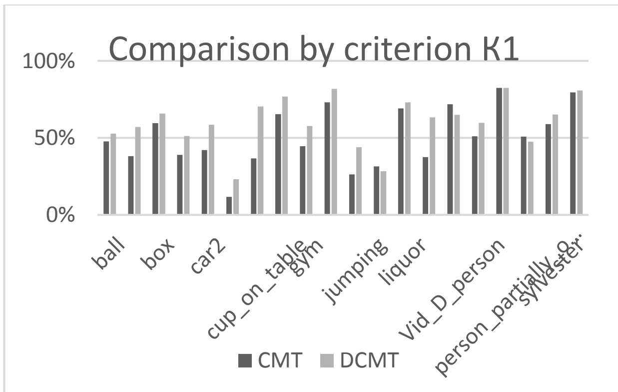

Figure 4a. Results of the algorithm in comparison with the original.

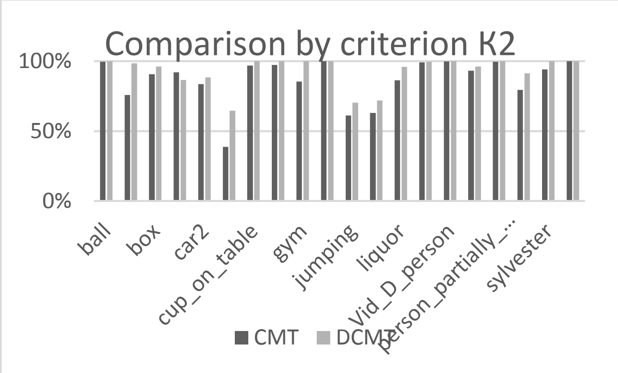

Figure 4b. Results of the algorithm in comparison with the original.

By the second criterion, the most optimal values are d = 0, k = 1. With these parameters, the average improvement was 3.45% for the first criterion, and for the second criterion it was 9.58%. The following results were obtained for different sequences (Fig. 5):

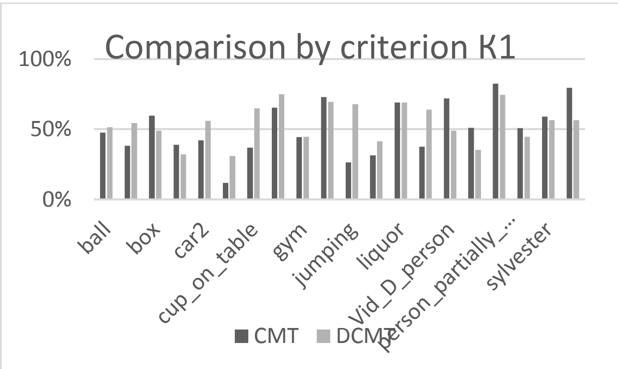

Figure 5a. Results of the algorithm in comparison with the original.

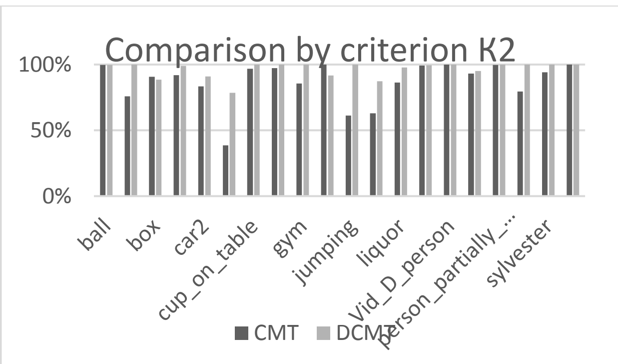

Figure 5b. Results of the algorithm in comparison with the original.

The best results on the first criterion could be achieved on the sequence "cup_on_table", the improvement was 33.6%. According to the second criterion, the best results on the "carchase" sequence, the improvement was 39.98%

5. CONCLUSION

In this work it we analyzed the existing development in the object tracking and our tracking algorithm was implemented using dynamically changing set of keypoints. In particular, the algorithm for rating the keypoints of the model has been proposed and implemented. Also considered is the construction of decisive rules for the addition of new points to the model based on FRiS

The influence of parameters on the quality of tracking was considered. It can be concluded that adding a small number of points to the model positively affects the accuracy of object tracking. Adding a large number of points to the model reduces the accuracy of the location but allows us detect the object on the frame more often.

The proposed changes have improved the accuracy of determining the location of the object by an average of 9.3%.

REFERENCES

Klein, D. A., Schulz, D., Frintrop, S., & Cremers, A. B., 2010, . Adaptive real-time video-tracking for arbitrary objects. In: *Intelligent Robots and Systems* (IROS), 2010 IEEE/RSJ International Conference on, pp. 772-777. IEEE.

Leutenegger, S., Chli, M., & Siegwart, R. Y., 2011. BRISK: Binary robust invariant scalable keypoints. In: *Computer Vision (ICCV)*, 2011 IEEE International Conference on, pp. 2548-2555. IEEE.

Lucas, B. D., & Kanade, T., 1981. An iterative image registration technique with an application to stereo vision.

Nebehay, G., & Pflugfelder, R., 2014. Consensus-based matching and tracking of keypoints for object tracking. InL *Applications of Computer Vision (WACV)*, 2014 IEEE Winter Conference on, pp. 862-869. IEEE.

Zagorujko N. G., 2013. Kognitivnyj analiz dannyh [Cognitive data analysis], Novosibirsk: Akademicheskoe izd-vo”GEO”