Real-time tracking of multiple objects with locally adaptive correlation filters

A.N. Ruchay [1], V.I. Kober [1], I.E. Chernoskulov [1]

[1] *Chelyabinsk State University, 129 Bratiev Kashirinykh st., Chelyabinsk 454001, Russia*

Abstract

A tracking algorithm using locally adaptive correlation filtering is proposed. The algorithm is designed to track multiple objects with invariance to pose, occlusion, clutter, and illumination variations. The algorithm employs a prediction scheme and composite correlation filters. The filters are synthesized with the help of an iterative algorithm, which optimizes discrimination capability for each target. The filters are adapted online to targets changes using information of current and past scene frames. Results obtained with the proposed algorithm using real-life scenes, are presented and compared with those obtained with state-of-the-art tracking methods in terms of detection efficiency, tracking accuracy, and speed of processing.

Keywords: tracking; locally adaptive filters; correlation filters; matching.

1. Introduction

Nowadays, object tracking is a widely investigated topic in engineering and computer vision [1, 2]. Video surveillance, vehicle navigation, human-computer interaction, and robotics are examples of tracking applications [3, 4, 5, 6, 7, 8, 9, 10, 11]. In tracking, objects are localized in a current frame automatically by applying a detection engine [12, 13, 14, 15]. A main difficulty in object tracking is that the observed scene is commonly degraded by additive noise, the presence of a cluttered background, geometric modifications such as pose changing and scaling, gesticulations, and nonuniform illumination. Additionally, eventual occlusions and real-time requirements are challenges that a modern tracking algorithm must solve.

Object tracking based on correlation-based methods are widely utilized as an attractive alternative to existing tracking algorithms [16, 17, 18]. Correlation filters have a good formal basis, and they can be easily implemented for real-time applications [19, 20]. Recognition methods involving template matching are not useful in some cases, for instance, when articulation changes global features like the object outline. So, conventional correlation filters without training may yield a poor performance to recognize objects possessing incomplete information [21, 22, 23]. Adaptive approach to the filter design helps us to synthesize adaptive filters for object tracking [24, 25].

In this work, we propose an algorithm for object tracking based on locally adaptive correlation filtering. The algorithm is able to carry out object tracking with a high accuracy in an video without offline training. The objects are selected at the beginning of the algorithm. Afterwards, a composite correlation filter optimized for distortion tolerant pattern recognition is designed to recognize the target in the next frame. The impulse responses of optimum correlation filters are used to synthesize composite filters for distortion invariant object tracking. Two techniques are used to improve the detection performance: adaptive procedure that achieves a prespecified performance for a typical scene background, and multiple composite filters (bank of composite filters) when numerous views are available for training. The filter is dynamically adapted to each frame using information of current and past scene observations.

The paper is organized as follows. Section 2 recalls the optimum composite filter design. Section 3 describes the suggested algorithm for object tracking by locally adaptive correlation filtering. Computer simulation results obtained with the proposed algorithm are presented and compared with common algorithms in terms of detection efficiency and location accuracy in section 4. Finally, section 5 presents our conclusions.

2. Composite filter design using optimum correlation filters

We are interested in the design of a correlation filter which is able to recognize an object embedded into a disjoint background in the scene corrupted with additive noise. The designed filter should be also able to recognize geometrically distorted versions of the target. Let $T = \{t_i(x, y); i = 1, \ldots, N\}$ be an image set containing geometrically distorted versions of the target to be recognized. The input scene is assumed to be composed by the target $t(x, y)$ embedded into a disjoint background $b(x, y)$ at unknown coordinates (τ_x, τ_y), and the whole scene is corrupted with additive noise $n(x, y)$, as follows:

$$f(x, y) = t(x - \tau_x, y - \tau_y) + b(x, y)\overline{w}(x - \tau_x, y - \tau_y) + n(x, y), \qquad (1)$$

where $\bar{w}(x, y)$ is a binary function defined as zero inside the target area, and unity elsewhere. The optimum filter for detecting the target, in terms of the signal to noise ratio (SNR) and the minimum variance of measurements of location errors (LE), is the generalized matched filter (GMF) [26], whose frequency response is given by

$$H^*(u, v) = \frac{T(u, v) + \mu_b \overline{W}(u, v)}{P_b(u, v) \otimes |\overline{W}(u, v)|^2 + P_n(u, v).} \qquad (2)$$

In (2), $T(u,v)$ and $\overline{W}(u,v)$ are the Fourier transforms of $t(x,y)$ and $\overline{w}(x,y)$, respectively; μ_b is the mean value of the background $b(x,y)$; $P_b(u,v)$ and $P_n(u,v)$ denote power spectral densities of $b_0(x,y) = b(x,y) - \mu_b$ and $n(x,y)$, respectively. The symbol $\otimes$ denotes convolution.

Let $h_i(x,y)$ be the impulse response of a GMF constructed for the ith available view of the target $t_i(x,y)$ in T. Let $H = \{h_i(x,y); i = 1,\ldots,N\}$ be the set of all GMF impulse responses constructed for all training images $t_i(x,y)$. Additionally, let $S = \{s_i(x,y); i = 1,\ldots,M\}$ be an image set containing M unwanted patterns to be rejected. We want to synthesize a filter capable to recognize all target views in T and to reject the false patterns in S, by combining the optimum filter templates contained in H, and by using only a single correlation operation. The required filter $p(x,y)$, can be constructed as follows [26]:

$$p(x,y) = \sum_{i=1}^{N} \alpha_i h_i(x,y) + \sum_{i=N+1}^{N+M} \alpha_i s_i(x,y), \tag{3}$$

where the coefficients $\{\alpha_i; i = 1,\ldots,N+M\}$ are chosen to satisfy prespecified output values for each pattern in $U = T \cup S$. Using vectormatrix notation, we denote by $\mathbf{R}$ a matrix with $N + M$ columns, where each column is the vector version of each element of U. Let $\mathbf{a} = [\alpha_i; i = 1,\ldots,N+M]^T$ be a vector of coefficients. Thus, (3) can be rewritten as

$$\mathbf{p} = \mathbf{R}\mathbf{a}. \tag{4}$$

Let us denote by

$$\mathbf{u} = \Bigg[\underbrace{1,\ldots,1}_{N\text{ones}},\underbrace{0,\ldots,0}_{M\text{zeros}}\Bigg]^T,$$

the desired responses to the training patterns, and denote by $\mathbf{Q}$ the matrix whose columns are the elements of U. The response constraints can be expressed as

$$\mathbf{u} = \mathbf{Q}^{+}\mathbf{p}, \tag{5}$$

where superscript $^{+}$ denotes complex conjugate. Substituting (4) into (5), we obtain

$$\mathbf{u} = \mathbf{Q}^{+}\mathbf{R}\mathbf{a}.$$

Thus, the solution for $\mathbf{a}$, is

$$\mathbf{a} = [\mathbf{Q}^{+}\mathbf{R}]^{-1}\mathbf{u}. \tag{6}$$

Finally, substituting (8) into (4), the solution for the composite filter is given by

$$\mathbf{p} = \mathbf{R}[\mathbf{Q}^{+}\mathbf{R}]^{-1}\mathbf{u}. \tag{7}$$

Note that the value of the correlation peak when using the filter given in Eq. 7, is expected to be close to unity for true-class objects, and close to zero for false-class objects.

The discrimination capability (DC) is a measure of the ability of the filter to distinguish a target from unwanted objects; it is defined by [26]

$$DC = 1 - \frac{|c^b|^2}{|c^t|^2},$$

where c^b is the value of the maximum correlation sidelobe in background area and c^t is the value of the correlation peak generated by the target. A DC value close to unity indicates that the filter has a good capability to distinguish between the target and any false object. Negatives values of the DC indicate that the filter is unable to detect the target. Also, if the obtained DC is greater than a prespecified threshold ($DC > DC_{th}$), then the target is considered as detected and, otherwise, the target is rejected.

3. Object tracking with locally adaptive correlation filtering

In this section we describe the proposed algorithm for object tracking based on composite correlation filtering. The proposed algorithm is robust to pose changes and appearance modifications of objects, as well as to the presence of scene noise, illumination changes, and target occlusions.

The algorithm starts with an initialization step where the objects are selected. Next, an optimum correlation filter for reliable detection and location estimation of the target is designed. Afterwards, a composite locally adaptive correlation filter is synthesized. The proposed algorithm incorporates an automatic re-initialization mechanism that reestablishes the tracking if it fails. The block diagram of the proposed algorithm is depicted in Fig. 1. The detailed operation steps are explained below.

Step 1: For each object select a small target $t_i(x,y)$ from a captured scene frame $f_i(x,y)$ containing the object to be tracked.

Step 2: Synthesize an optimum correlation filter $h_i(x,y)$ with (2) for reliable detection and location estimation of the target $t_i(x,y)$ in the observed local frame $l_i(x,y)$.

Step 3: Synthesize a composite locally adaptive correlation filter $p_i(x, y)$ as follows. First, detect and locate the target by $h_i(x, y)$ filter from the observed local frame $l_i(x, y)$. If the obtained DC is greater than a prespecified threshold ($DC > DC_{rec}$), then the target is considered as successfully detected, $t_i(x, y)$ added into the set T and recursion should be stopped. Otherwise, the target $s_i(x, y)$ corresponding to a false peak added into the set S. Second, synthesize a composite filter $p_i(x, y)$ with the help of (7). Third, detect and locate the target by $p_i(x, y)$ filter from the observed local frame $l_i(x, y)$ recursively until the condition $DC > DC_{rec}$ is satisfied.

Step 4: Detect and locate the target in the observed local frame $l_{i+1}(x, y)$ from a new scene frame $f_{i+1}(x, y)$ by $p_i(x, y)$ filter. The coordinates of the observed local frame $l_{i+1}(x, y)$ are provided by a prediction process that analyzes the motion kinematics of the target. If the obtained DC is greater than a prespecified threshold ($DC > DC_{th}$), then the target is considered as successfully detected and $p_i(x, y)$ filter added to the bank B of composite correlation filters. Otherwise, the target is lost in the observed local frame $l_{i+1}(x, y)$ and we recursively used the filters from bank B until condition $DC > DC_{con}$ is satisfied. The filter from bank B with condition $DC > DC_{con}$ is used to a new scene frame. If the target is lost in the observed local frame $l_{i+1}(x, y)$ with help the filters from bank B, then the coordinates of the target is set coordinates of the past scene frame $f_i(x, y)$ and we proceed to a new scene frame $f_{i+2}(x, y)$.

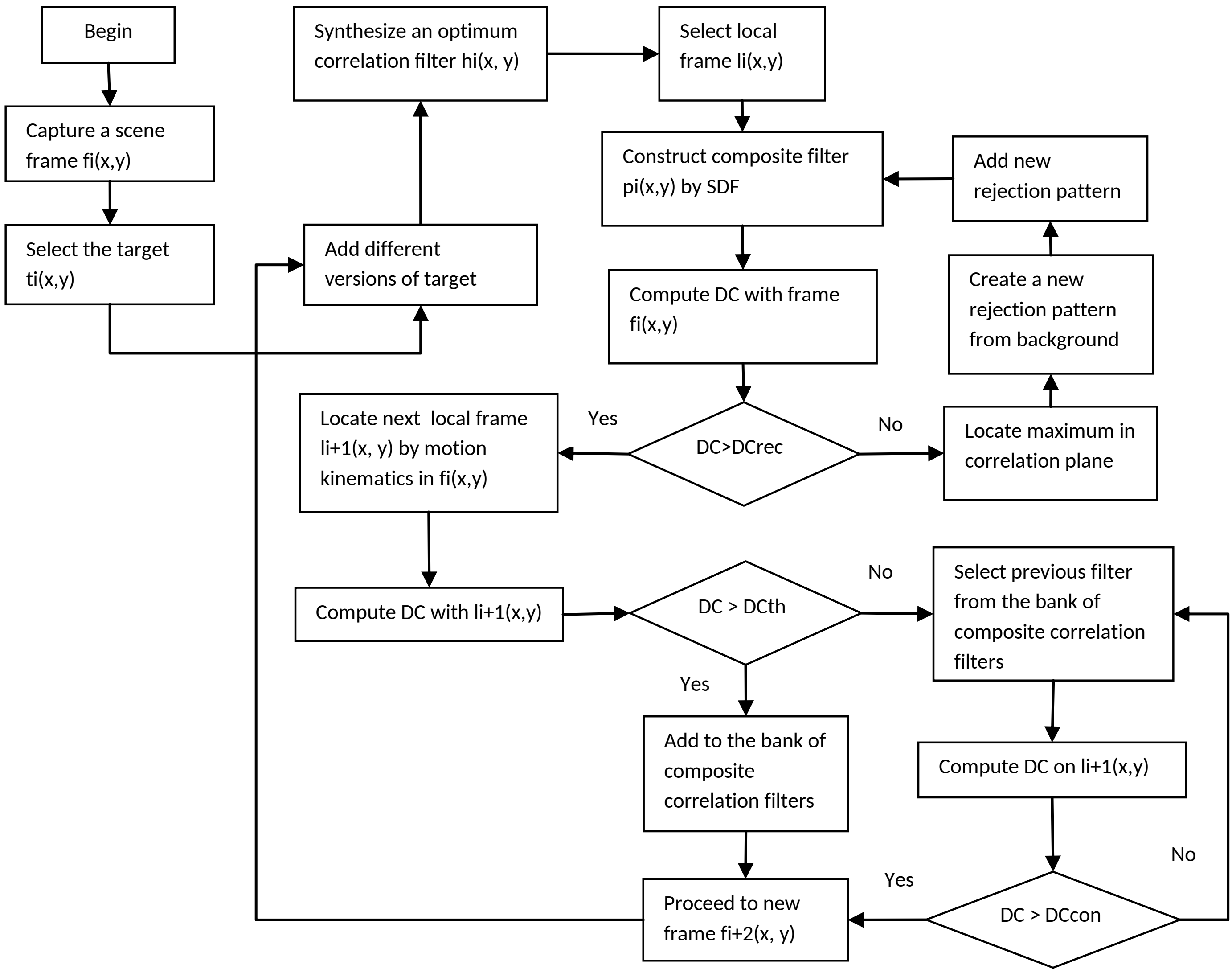

Fig. 1. Block diagram of the proposed tracking algorithm based on locally adaptive correlation filtering.

4. Computer simulation

In this section, computer simulation results obtained with the proposed algorithm for object tracking are presented and compared with common algorithms in terms of detection efficiency, tracking accuracy, and speed of processing.

In order to evaluate the performance of our tracker, we conduct experiments on 100 challenging image sequences from Object Tracking Benchmark (TB-100 database) [27]. These sequences cover most challenging situations in object tracking: Illumination Variation (IV), Scale Variation (SV), Occlusion (OCC), Deformation (DEF), Motion Blur (MB), Fast Motion (FM), In-Plane Rotation (IPR), Out-of-Plane Rotation (OPR), Out-of-View (OV), Background Clutters (BC), Low Resolution (LR).

For comparison, we run 3 state-of-the-art algorithms with the same initial position of the target. The first tracking algorithm (SURF) [28] is based on matching of local features and descriptors. The second tracking algorithm (STRUCK) predicts the target location change between frames on the basis of structured learning [29]. The third collaborative tracking algorithm (SCM) is combined a sparsity-based discriminative classifier and a sparsity-based generative model [30]. The work [27] performed large-scale experiments to evaluate the performance of recent 33 object-tracking algorithms. Tracking algorithms STRUCK and SCM perform much better than the others.

For evaluating of detection efficiency we use an evaluation metric of the overlap score. Given a tracked bounding box r_t and the ground-truth bounding extent r_0 of a target object, the overlap score is defined as

$$S = \frac{\|r_t \cap r_0\|}{\|r_t \cup r_0\|}, \tag{8}$$

where $\cap$ and $\cup$ represent the intersection and union operators, respectively, and $\|\cdot\|$ denotes the number of pixels in a region. This average overlap score (AOS) can be used as the performance measure. In addition, the overlap scores can be used for determining whether an algorithm successfully tracks a target in a frame, by testing whether S is larger than a threshold of 0.5. Also we evaluate the tracking algorithms using the average center location error (ACLE) for all image sequences from database.

Table 1 shows the average overlap score (AOS), the average center location errors (ACLE) and the Average Processing Time (APT) on a scena for all the tracking algorithms with the overlap threshold of 0.5. The evaluation results show that our proposed algorithm is faster than the others and more accurate in terms of the average center location errors.

Table 1. Evaluation results of the state-of-the-art STRUCK, SCM, SURF and proposed algorithms by the average overlap score (AOS), the average center location errors (ACLE), and the Average Processing Time (APT)

Tracker	All	BC	DEF	FM	IPR	IV	LR	MB	OCC	OPR	OV	SV	APT	ACLE
Proposed	53.3	50.7	51.1	60.0	56.4	43.5	56.7	55.7	44.6	50.5	41.7	51.4	0.2005	68.8
STRUCK	57.5	59.3	52.4	55.6	57.0	59.0	59.1	59.9	55.9	57.3	58.9	57.8	0.2894	61.5
SCM	54.4	61.3	51.5	42.8	51.8	61.1	61.7	45.2	56.8	57.0	56.4	55.8	0.3122	64.8
SURF	35.2	37.4	25.8	41.6	39.7	37.3	23.0	45.4	36.0	34.8	46,.7	33.0	0.1668	276.6

When an object moves fastly on the FM subset, the proposed algorithm performs much better than the others. However, the proposed algorithm does not perform well in the subset (IV, OCC, OV) due to illumination variation, and partial occlusion of the target. On the other subsets, the Struck, SCM, and the proposed algorithms outperform other the state-of-the-art algorithms. Fig. 2 shows sample tracking results of the proposed algorithms where the target objects are marked with red rectangles and the actually tracked objects by the proposed algorithm are marked with green rectangles.

Fig. 2. Results of tracking by proposed algorithm.

5. Conclusion

A tracking algorithm using locally adaptive correlation filtering is proposed. The algorithm is designed to track multiple objects with invariance to pose, partial occlusion, clutter, and illumination variations. The algorithm employs a prediction scheme and composite correlation filters. The filters are synthesized with the help of an iterative algorithm, which optimizes discrimination capability for each target. The filters are adapted online to targets changes using information of current and past scene frames. The evaluation results show that our proposed algorithm is faster than the others and more accurate in terms of the average center location errors. On the majority test sets the proposed algorithm performs much better than the state-of-the-art algorithms.

Acknowledgments

This work was supported by the Russian Science Foundation, grant no. 15-19-10010.

References

[1] Karasulu, B. Performance Evaluation Software: Moving Object Detection and Tracking in Videos [Text] / B. Karasulu, S. Korukoglu. ”— New York : Springer, 2013.

[2] Talmale, S. Object tracking in images and videos [Text] / S.K. Talmale, N.J. Janwe // International Journal Of Engineering And Computer Science. ”—2016. ”—Vol. 5(1). ”—P. 15482–15486.

[3] Accurate three-dimensional pose recognition from monocular images using template matched filtering [Text] / Kenia Picos, Victor H. Diaz-Ramirez, Vitaly Kober [et al.] // Optical Engineering. ”—2016. ”—Vol. 55, no. 6. ”—P. 063102.

[4] Echeagaray-Patron, B. A. Conformal parameterization and curvature analysis for 3d facial recognition [Text] / B. A. Echeagaray-Patron, D. Miramontes-Jaramillo, V. Kober // 2015 International Conference on Computational Science and Computational Intelligence (CSCI). ”— [S. l. : s. n.], 2015. ”—P. 843–844.

[5] Echeagaray-Patron, B. A. 3d face recognition based on matching of facial surfaces [Text] / Beatriz A. Echeagaray-Patron, Vitaly Kober. ”—Vol. 9598. ”—[S. l. : s. n.], 2015. ”—P. 95980V–95980V–8.

[6] Diaz-Escobar, J. A robust hog-based descriptor for pattern recognition [Text] / Julia Diaz-Escobar, Vitaly Kober. ”—Vol. 9971. ”—[S. l. : s. n.], 2016. ”—P. 99712A–99712A–7.

[7] Diaz-Escobar, J. Text Detection in Digital Images Captured with Low Resolution Under Nonuniform Illumination Conditions [Text] / Julia Diaz-Escobar, Vitaly Kober // Pattern Recognition: 8th Mexican Conference, MCPR 2016, Guanajuato, Mexico, June 22-25, 2016. Proceedings / Ed. by José Francisco Martínez-Trinidad, Jesús Ariel Carrasco-Ochoa, Victor Ayala Ramirez [et al.]. ”—Cham : Springer International Publishing, 2016. ”—P. 3–12.

[8] An efficient algorithm for matching of slam video sequences [Text] / Jose A. Gonzalez-Fraga, Victor H. Diaz-Ramirez, Vitaly Kober [et al.]. ”—Vol. 9971. ”—[S. l. : s. n.], 2016. ”—P. 99712Z–99712Z–10.

[9] Effective indexing for face recognition [Text] / I. Sochenkov, A. Sochenkova, A. Vokhmintsev [et al.]. ”—Vol. 9971. ”—[S. l. : s. n.], 2016. ”—P. 997124–997124–9.

[10] Face recognition based on a matching algorithm with recursive calculation of oriented gradient histograms [Text] / A. V. Vokhmintcev, I. V. Sochenkov, V. V. Kuznetsov, D. V. Tikhonkikh // Doklady Mathematics. ”—2016. ”—Vol. 93, no. 1. ”—P. 37–41.

[11] Tihonkih, D. A modified iterative closest point algorithm for shape registration [Text] / Dmitrii Tihonkih, Artyom Makovetskii, Vladislav Kuznetsov. ”—Vol. 9971. ”—[S. l. : s. n.], 2016. ”—P. 99712D–99712D–8.

[12] Miramontes-Jaramillo, D. A Robust Tracking Algorithm Based on HOGs Descriptor [Text] / Daniel Miramontes-Jaramillo, Vitaly Kober, Víctor Hugo Díaz-Ram // Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 19th Iberoamerican Congress, CIARP 2014, Puerto Vallarta, Mexico, November 2-5, 2014. Proceedings / Ed. by Eduardo Bayro-Corrochano, Edwin Hancock. ”—Cham : Springer International Publishing, 2014. ”—P. 54–61.

[13] Miramontes-Jaramillo, D. Multiple objects tracking with hogs matching in circular windows [Text] / Daniel Miramontes-Jaramillo, Vitaly Kober, Victor H. Diaz-Ramirez. ”—Vol. 9217. ”—[S. l. : s. n.], 2014. ”—P. 92171N–92171N–8.

[14] Miramontes-Jaramillo, D. Robust illumination-invariant tracking algorithm based on hogs [Text] / Daniel Miramontes-Jaramillo, Vitaly Kober, Víctor Hugo Díaz-Ramírez. ”—Vol. 9599. ”—[S. l. : s. n.], 2015. ”—P. 95991Q–95991Q–8.

[15] Miramontes-Jaramillo, D. Real-time tracking based on rotation-invariant descriptors [Text] / Daniel Miramontes-Jaramillo, Vitaly Kober // 2015 International Conference on Computational Science and Computational Intelligence (CSCI). ”—2015. ”—Vol. 00. ”—P. 543–546.

[16] Ontiveros-Gallardo, S. E. Objects tracking with adaptive correlation filters and kalman filtering [Text] / Sergio E. Ontiveros-Gallardo, Vitaly Kober. ”—Vol. 9598. ”—[S. l. : s. n.], 2015. ”—P. 95980X–95980X–8.

[17] Ontiveros-Gallardo, S. E. Correlation-based tracking using tunable training and kalman prediction [Text] / Sergio E. Ontiveros-Gallardo, Vitaly Kober. ”—Vol. 9971. ”—[S. l. : s. n.], 2016. ”—P. 997129–997129–9.

[18] Ruchay, A. A correlation-based algorithm for recognition and tracking of partially occluded objects [Text] / Alexey Ruchay, Vitaly Kober. ”—Vol. 9971. ”—[S. l. : s. n.], 2016. ”—P. 99712R–99712R–9.

[19] Facial recognition using composite correlation filters designed with multiobjective combinatorial optimization [Text] / Andres Cuevas, Victor H. Diaz-Ramirez, Vitaly Kober, Leonardo Trujillo. ”—Vol. 9217. ”—[S. l. : s. n.], 2014. ”—P. 921710–921710–8.

[20] Aguilar-González, P. M. Adaptive composite filters for pattern recognition in nonoverlapping scenes using noisy training images [Text] / Pablo Mario Aguilar-González, Vitaly Kober, Víctor Hugo Díaz-Ramírez // Pattern Recogn. Lett. ”—2014. ”—Vol. 41. ”—P. 83–92.

[21] Díaz-Ramírez, V. H. Object Tracking in Nonuniform Illumination Using Space-Variant Correlation Filters [Text] / Víctor Hugo Díaz-Ramírez, Kenia Picos, Vitaly Kober // Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 18th Iberoamerican Congress, CIARP 2013, Havana, Cuba, November 20-23, 2013, Proceedings, Part II / Ed. by José Ruiz-Shulcloper, Gabriella Sanniti di Baja. ”— Berlin, Heidelberg : Springer Berlin Heidelberg, 2013. ”—P. 455–462.

[22] Real-time tracking of multiple objects using adaptive correlation filters with complex constraints [Text] / Victor H. Diaz-Ramirez, Viridiana Contreras, Vitaly Kober, Kenia Picos // Optics Communications. ”—2013. ”—Vol. 309. ”—P. 265—278.

[23] Diaz-Ramirez, V. H. Target tracking in nonuniform illumination conditions using locally adaptive correlation filters [Text] / Victor H. Diaz-Ramirez, Kenia Picos, Vitaly Kober // Optics Communications. ”—2014. ”—Vol. 323. ”—P. 32—43.

[24] Robust Face Tracking with Locally-Adaptive Correlation Filtering [Text] / Leopoldo N. Gaxiola, Víctor Hugo Díaz-Ramírez, Juan J. Tapia [et al.] // Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 19th Iberoamerican Congress, CIARP 2014, Puerto Vallarta, Mexico, November 2-5, 2014. Proceedings / Ed. by Eduardo Bayro-Corrochano, Edwin Hancock. ”—Cham : Springer International Publishing, 2014. ”—P. 925–932.

[25] Target tracking with dynamically adaptive correlation [Text] / Leopoldo N. Gaxiola, Victor H. Diaz-Ramirez, Juan J. Tapia, Pascuala Garcia-Martinez // Optics Communications. ”—2016. ”—Vol. 365. ”—P. 140 – 149.

[26] Ramos-Michel, E. M. Adaptive composite filters for pattern recognition in linearly degraded and noisy scenes [Text] / Erika M. Ramos-Michel, Vitaly Kober // Optical Engineering. ”—2008. ”—Vol. 47, no. 4. ”—P. 047204–047204–7.

[27] Wu, Y. Object tracking benchmark [Text] / Y. Wu, J. Lim, M. H. Yang // IEEE Transactions on Pattern Analysis and Machine Intelligence. ”— 2015. ”—Vol. 37, no. 9. ”—P. 1834–1848.

[28] Al-asadi, T. Object detection and recognition by using enhanced speeded up robust feature [Text] / T.A. Al-asadi, A.J. Obaid // International Journal of Computer Science and Network Security. ”—2016. ”—Vol. 16(4). ”—P. 66–71.

[29] Torr, P. H. S. Struck: Structured output tracking with kernels [Text] / Philip H. S. Torr, Sam Hare, Amir Saffari // 2011 IEEE International Conference on Computer Vision (ICCV 2011). ”—2011. ”—Vol. 00. ”—P. 263–270.

[30] Zhong, W. Robust object tracking via sparsity-based collaborative model [Text] / Wei Zhong // Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). ”—CVPR ’12. ”—Washington, DC, USA : IEEE Computer Society, 2012. ”—P. 1838–1845.