

## PUPIL VISUAL TRACKING ALGORITHMS FOR AUTOMATED STATIC PERIMETRY SYSTEMS

P. N. Skonnikov<sup>1,2\*</sup>, D. V. Trofimov<sup>1</sup>

<sup>1</sup> RIF Corporation, Voronezh, Russia – samael1978@rambler.ru

<sup>2</sup> Ryazan State Radio Engineering University named after V.F. Utkin, Ryazan, Russia – skonnikovpn@yandex.ru

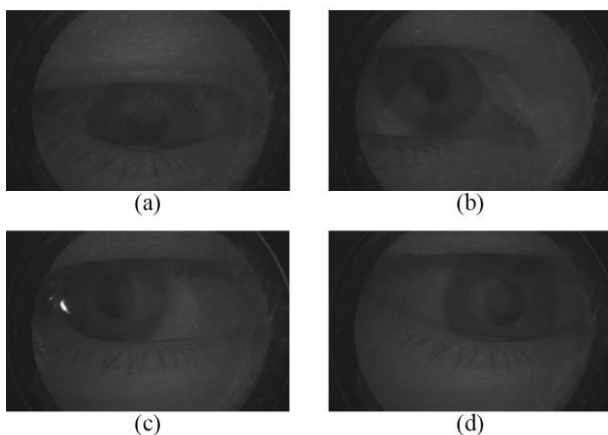
**KEY WORDS:** Static Perimetry, Segmentation, Tracking, Optical Flow, Weighted Average.

### ABSTRACT:

Some diseases, for instance, a glaucoma, cause visual field defects. For the timely diagnostics of such defects, various methods are used. One of the state-of-the-art diagnostic methods is automated static perimetry. The method of static perimetry consists in the light sensitivity determination in different parts of the visual field using stationary objects of variable luminosity. When scanning the visual field in this way, an important factor is the control of gaze fixation at the fixation point. The greatest accuracy in determining the gaze fixation position is achieved by the method of the pupil visual tracking using a video camera. In this paper, four groups of visual tracking algorithms are considered: segmentation-based methods, correlation methods, methods based on optical flow and on weighted average. An experimental comparison of these methods was carried out using the base of video recordings obtained in the automatic static perimetry apparatus. On these videos the ground truth tracks of pupil were marked. The comparison was conducted according to two criteria: center location error and tracking length. It is shown that only the weighted average method has an acceptable tracking length.

### 1. INTRODUCTION

One of the state-of-the-art methods of medical examination of visual fields is automated static perimetry (Delgado et al, 2002). The essence of this examination is that the patient is presented with luminous point stimuli in different zones of the visual field. Evaluation is whether the patient sees stimuli in all areas of the visual field. In order to know which position of the visual field the stimulus locates, it is necessary to have information about the gaze direction and to establish the fact of its fixation at the fixation point. It is possible to control the gaze fixation through video recording of the eye using a video camera installed inside the automatic static perimetry apparatus. The typical images of human eye filmed by used automated static perimetry apparatus are shown on Figure 1.



**Figure 1.** The typical images of human eye filmed by used automated static perimetry apparatus.

In order to determine the gaze direction, methods of pupil tracking on video are used. At present, various tracking

algorithms are known, and the issues of their use in tasks of automatic static perimetry have been little studied. Hence, the issue of selection the pupil tracking algorithm for automated static perimetry systems is relevant.

In this paper, four groups of visual tracking algorithms are considered: segmentation-based methods, correlation methods, methods based on optical flow and on weighted average.

### 2. USED METHODS

#### 2.1 Segmentation-based methods

Segmentation-based methods generate a label matrix from the eye image (Gonzalez and Woods, 2008). To improve the segmentation quality, the images were preprocessed. Preprocessing is performed in each frame and consists in increasing the contrast and smoothing. First, the 5th and 95th percentiles of source image  $I_0$  were calculated:  $\alpha_5$  and  $\alpha_{95}$ , correspondingly. Preprocessed image  $I_C$  is defined using linear transform:

$$I_C = 255 \frac{I_0 - \alpha_5}{\alpha_{95}}. \quad (1)$$

Then the image is smoothed. To obtain a smoothed image  $I_S$ , the resulting image is filtered with a square linear filter. In further experiments, a filter size of  $3 \times 3$  elements was used:

$$I_S = I_C * \frac{1}{9} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}. \quad (2)$$

\* Corresponding author

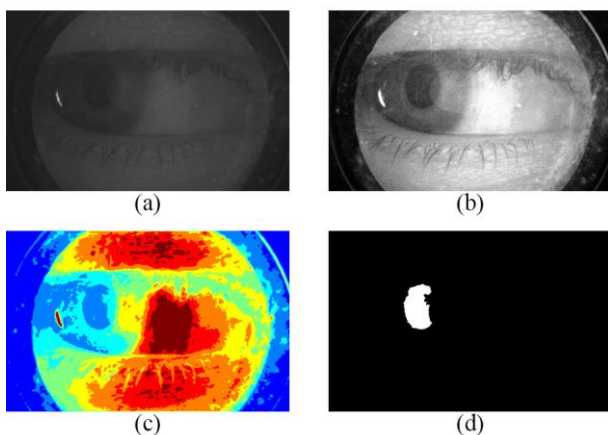
A large number of segmentation methods are known, the survey on image segmentation techniques is carried out by Zaitoun and Aqel (2015).

In this paper, the main ideas of multi-threshold segmentation described by Babayan, Skonnikov and Trofimov (2020), are realized by Otsu method (Otsu, 1979).

The label matrix obtained as a result of segmentation is divided into binary masks. These binary images are processed using morphological methods based on erosion and dilation: cleaning and closing (Gonzalez, Woods and Eddins, 2020).

Connected regions are selected on the processed binary masks. A set of parameters is calculated for each region: centroid coordinates, area, width, height, perimeter length. Then the regions are filtered: regions with atypical parameter values are discarded. One of the remaining regions with the smallest perimeter-area ratio, i.e., the one most similar to a circle by this criterion, is selected. The centroid coordinates of the selected region are considered the coordinates of the pupil.

Figure 2 shows as an example the main stages of processing by the segmentation method.



**Figure 2.** The main stages of methods based on segmentation: (a) – original image, (b) – preprocessed image, (c) – label matrix, (d) – binary mask after morphological processing.

Analysis of this figure shows, that as a result of processing, the object corresponding to the pupil is correctly identified. In the example shown, the pupil is not fully segmented due to the effect of the flare. Therefore, the selected object has a non-elliptical shape. Besides that, other segments of similar area are also present on the label matrix.

## 2.2 Correlation methods

Correlation methods are based on comparing different parts of the source image with a predefined reference. The general principles of correlation algorithms can be described as follows (Yilmaz, Javed and Shah, 2006).

There is a reference  $\mathbf{R}$  with the size of  $S_{RX} \times S_{RY}$  elements. This reference moves through the search area  $\mathbf{I}$  of  $S_{IX} \times S_{IY}$  elements with shifts of  $x = \overline{1, S_X}$  and  $y = \overline{1, S_Y}$  along two coordinates, where the maximum shift values are  $S_X = S_{RX} - S_{IX} + 1$  and  $S_Y = S_{RY} - S_{IY} + 1$ . For each shift value, a measure of similarity (or difference)  $F$  between the reference and the search area corresponding to the current values  $x$  and  $y$  is determined. The dependence of this measure on the values of the shift  $F(x, y)$  is called the criterion function.

The comparison criterion can be different. In this paper, the normalized sum of the absolute values of the difference between the reference and the source image region was chosen as a criterion.

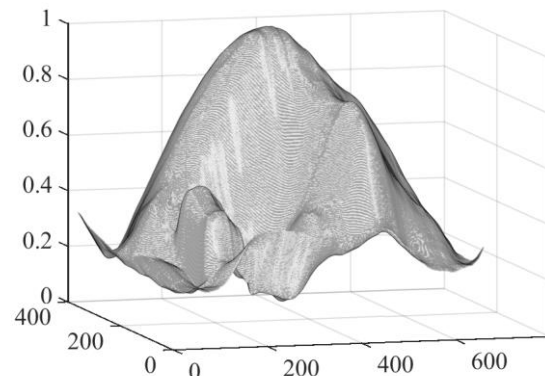
$$F(x, y) = \sum_{i=1}^{S_{RX}} \sum_{j=1}^{S_{RY}} |I_{i+x-1, j+y-1} - R_{i,j}|, \quad (3)$$

where  $I$  and  $R$  are the elements of the search area and the reference, respectively.

The different positions of the reference on the source image correspond to the criterion function values. The coordinates of the pupil  $x_p$  and  $y_p$  are determined by position of the criterion function minimum:

$$\{x_p, y_p\} = \arg \left( \min_{x,y} (F(x, y)) \right). \quad (4)$$

In correlation tracking methods, the reference can be captured manually in the first frame or set a priori. In order to automate the tracking process several references were set a priori. The pupil position was determined using the one reference appropriate the lowest value of the criterion function minimum. Figure 3 shows a criterion function typical for pupil tracking in the automatic perimetry apparatus.



**Figure 3.** Example of the criterion function typical for pupil tracking in the automatic perimetry apparatus

This figure shows that the criterion function contains extra minima in addition to the minimum corresponding to the true position of the pupil.

## 2.3 Optical flow

One of the frequently used groups of tracking methods are methods based on the optical flow calculation (Zhao, Shi, Chen, Li and Wang, 2015).

The method based on optical flow was implemented as follows. First, the optical flow was calculated for the current frame. In this paper, three methods have been tested to calculate the dense optical flow: Lucas-Kanade derivative of Gaussian method, Horn-Schunck method (Barron, Fleet, Beauchemin, Burkitt, 1992) and Farneback method (Farneback, 2003).

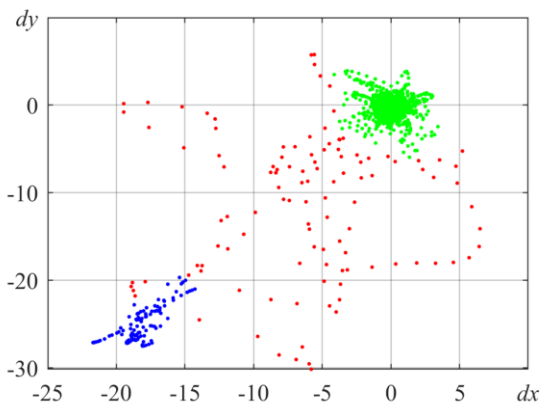
Preliminary studies have shown that the results obtained by the Farneback method are well grouped into two separate clusters – the object and the background, while the other methods provide too fuzzy separation between the clusters. Therefore, this method is hereinafter referred as the calculation of the optical flow.

The result of calculating the optical flow can be represented as a data set consisting of displacements in two coordinates  $dx$  and  $dy$ . The obtained two-dimensional set of displacements was

clustered. The k-means (Arthur, Vassilvitskii, 2006) and DBSCAN (Ester, Kriegel, Sander, Xiaowei, 1996) clustering algorithms were tested.

According to the results of the studies given below, the best data clustering was achieved using the DBSCAN method.

Figure 4 shows the result of points clustering, representing a displacements in two coordinates for different pixels using the DBSCAN method. For clarity of displaying a large number of points in this figure, the data was decimated. Here, the points recognized as noise are displayed in red, and points belonging to the two found clusters are displayed in blue and green.



**Figure 4.** The result of clustering the displacements obtained by calculating the optical flow

The one of the obtained clusters corresponding to the pupil was selected. Several criteria were used to determine which cluster corresponds to the pupil and which to the background. These include the number of cluster elements  $N$  (that is, the area of the segment in image coordinates) and the average speed of movement in the cluster  $\bar{v}$ . The number of elements of the cluster marked in green in Figure 3 is more than 80% of all image pixels, and the average velocities within this cluster in both coordinates are close to zero. According to the specified criteria, it was determined that this cluster  $C_0$  corresponds to the background. The only remaining cluster  $C_p$ , that is, the cluster marked in blue, was assigned to the pupil.

Pupil position was determined using two ways. The first way is to calculate the centroid coordinates of the selected cluster, as it was done in the tracking method based on brightness segmentation.

The second way is to add the average displacement of the current cluster to the pupil coordinates from the last frame.

It should be noted that the second method is characterized by the accumulation of errors.

In this work, if there is a jerky movement in the frame, the first method was used – the calculation of the average position of pixels belonging to the cluster  $C_p$ :

$$x_p = \frac{1}{N(C_p)} \sum_{C_p} x_n; \quad y_p = \frac{1}{N(C_p)} \sum_{C_p} y_n, \quad (5)$$

where  $x_n$  and  $y_n$  – coordinates of the pixel with the ordinal number  $n$ .

A jerky movement  $\bar{v}$  was determined by the excess of the average speed of this cluster of a certain empirically selected threshold  $\lambda$ .

If there is no sharp movement  $\bar{v} \leq \lambda$  in the frame number  $n_f$  or there is no cluster  $C_p$  assigned to the pupil, the second method

was used: correction of the previously determined position of the pupil by the displacements  $\bar{dx}$  and  $\bar{dy}$  obtained during the calculation of the optical flow:

$$\begin{aligned} x_p(n_f) &= x_p(n_f - 1) + \bar{dx}(n_f); \\ y_p(n_f) &= y_p(n_f - 1) + \bar{dy}(n_f). \end{aligned} \quad (6)$$

The values of  $\bar{dx}$  and  $\bar{dy}$  were determined similarly to formula (5) by averaging the elements of dense optical flow  $dx_n$  and  $dy_n$  over region  $C_p$ , and the average velocity  $\bar{v}$  was determined as

$$\bar{v} = \sqrt{(\bar{dx})^2 + (\bar{dy})^2}. \quad (7)$$

In cases where the cluster  $C_p$  was not determined, averaging was performed over the entire image area.

## 2.4 Weighted average

The fourth method is based on weighted average values (van Assen, Egmont-Petersen, Reiber, 2002).

First, a region of interest  $ROI$  is allocated, which is an inverted fragment of the original image  $I$  with borders  $x_{min}$ ,  $x_{max}$ ,  $y_{min}$ ,  $y_{max}$ :

$$ROI(i, j) = 255 - I(x_{min} + i - 1, y_{min} + j - 1), \quad (8)$$

where  $i = \overline{1, S_x}$ ,  $j = \overline{1, S_y}$ , and size of the region of interest is  $S_x = x_{max} - x_{min} + 1$ ,  $S_y = y_{max} - y_{min} + 1$ .

For this region of interest, the mean values of the coordinate grid, weighted by the negative of the source image, are calculated:

$$\begin{aligned} p &= x_{min} + \sum_{i=1}^{S_x} \sum_{j=1}^{S_y} (i \cdot w(i, j)); \\ q &= y_{min} + \sum_{i=1}^{S_x} \sum_{j=1}^{S_y} (j \cdot w(i, j)). \end{aligned} \quad (9)$$

Weights  $w$  are the normalized pixel brightness of the region of interest:

$$w(i, j) = ROI(i, j) / \sum_{i=1}^{S_x} \sum_{j=1}^{S_y} ROI(i, j). \quad (10)$$

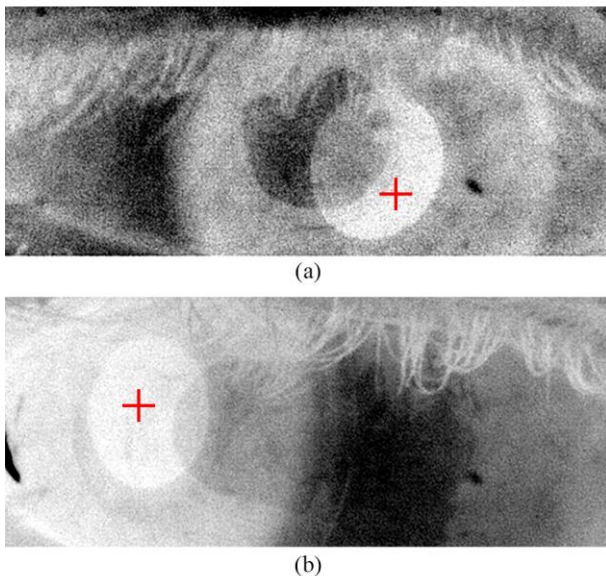
The values  $p$  and  $q$  themselves are not estimates of pupil position  $x_p$  and  $y_p$ . In order to determine the position of the pupil, an additional transformation must be performed. In the simplest case, the dependences  $x_p(p)$  and  $y_p(q)$  have a linear form:

$$x_p(p) = k_x p + b_x; \quad y_p(q) = k_y q + b_y. \quad (11)$$

The adjustment factors  $k_x$  and  $b_x$  are determined experimentally during calibration. These values can be found by solving a system of two equations, although for a more accurate calculation it is advisable to use methods that allow for a larger number of points, for example, RANSAC (Fischler, Bolles, 1981):

$$\begin{cases} x_{p1} = k_x p_1 + b_x; \\ x_{p2} = k_x p_2 + b_x. \end{cases} \quad (12)$$

The values of the coefficients  $k_y$  and  $b_y$  are determined in a similar way. The displacement correction of the region of interest  $(x_{min}, y_{min})$  relative to the border of the image  $I$  may not be taken into account separately if calculations by formulas (11) and (12) are performed in the coordinates of the original image, and not in the coordinates of the region of interest. Figures 5 (a) and (b) show two examples of different inverted regions of interest, in which the obtained pupil center coordinates estimation results are marked with cross markers. In this figure, for clarity, contrast was performed according to formula (1), although this operation is not required to determine the center of the pupil.



**Figure 5.** Determination of the pupil position by the method of weighted average

In this figure, it is shown that the marker position, determined by the weighted average method, is not the exact center of the pupil, but is located within the pupil.

### 3. EXPERIMENTAL RESULTS

The described algorithms were tested on eight video recordings of the human eye in an automated static perimetry apparatus, each 600 frames long.

Ground truth tracks of the pupil are marked on each of the eight videos in semi-automatic mode.

Taking into account the tracking quality criteria described by Dutta et al. (2019), two pupil tracking quality assessments were calculated; mean center location error  $\delta$  during successful tracking (prior to tracking failure) and mean tracking length  $n_{f \max}$ .

Mean center location error  $\delta$  was calculated as average Euclidian distance  $r$  between ground truth position of pupil center  $x_{GT}, y_{GT}$  and the corresponding tracked value  $x_p$  and  $y_p$ :

$$\delta = \frac{1}{n_{f \max}} \sum_{n_f=1}^{n_{f \max}} r(n_f), \quad (13)$$

where the distance  $r(n_f)$  at the frame number  $n_f$  was calculated as follows:

$$r(n_f) = \sqrt{(x_p - x_{GT})^2 + (y_p - y_{GT})^2}. \quad (14)$$

The situations when the position of the pupil center was not established or the current error exceeded the typical radius of the pupil of 50 pixels was taken as tracking failure. The experimental studies result of the four considered pupil tracking algorithms quality are shown in Table 1.

Method	Mean center location error $\delta$ , pixels	Mean tracking length $n_{f \max}$ , frames
Segmentation	4.7	27.375
Correlation	11.4	193
Optical flow	16.1	58.25
Weighted average	19.6	545.875*

**Table 1.** The experimental studies result of the four considered pupil tracking algorithms quality.

Here, the \* symbol for the weighted average method means that in six experiments out of eight, no tracking failure occurred during the entire video duration of 600 frames. In these cases, it was accepted  $n_{f \max} = 600$ . With other tracking methods, no video has achieved this tracking length.

From the analysis of this table, it follows that the weighted average method has the lowest accuracy in determining the center of the pupil among the methods considered, but this method is resistant to tracking failures, which is confirmed by the relatively high values of the tracking length.

### 4. CONCLUSIONS

Based on the experiment results, the following conclusions are drawn.

The segmentation-based method allows obtaining fairly accurate pupil center coordinates in cases where the pupil mask was extracted correctly. However, error often occurs at any of several steps in a segmentation-based algorithm. As a result, the region corresponding to the pupil is incorrectly determined. In this regard, most of the results obtained by segmentation methods are incorrect.

Correlation methods also often give a large error. There are two reasons for this. First, the real images of the human eye significantly differ from the references chosen a priori. An increase in the number of references leads to an unacceptable increase in computational capability. Secondly, the eye appearance changes significantly during the checkup process, and these changes lead to the tracking failure.

Methods based on optical flow provide high accuracy in measuring the pupil coordinates, but only under two conditions. Firstly, the optical flow determination is possible only at the moments of eye movement; this method is not applicable to static video. Secondly, facial movements introduce a significant error, and strict face fixation is difficult in real conditions. In this regard, this method can only be considered as an addition to another one.

The method based on the weighted average calculation provides the most stable results, but the measurement accuracy does not allow to estimate the accurate position of the pupil. At the same time, for the tasks of automated static perimetry, the information that the gaze is fixed in a certain zone is enough.

This can be explained as follows. The beginning of the gaze movement indicates that a new stimulus is visible and the eye begins to change its direction. After stopping the eye movement, it can be assumed that the gaze is fixed at the point corresponding to the stimulus, since there are no other objects in the field of view, and when the current stimulus was appeared, the eye began to move in its direction. If the gaze does not move in the appropriate direction when the stimulus position is changed, then we can conclude that the visual field has a defect in this place. The starting of eye movement indicates a normal field of view at corresponding point.

Thus, according to the results of experimental studies, it was found that the method based on the weighted average is best suited for pupil visual tracking in automated static perimetry systems.

## REFERENCES

- Arthur, D., Vassilvitskii, S., 2006: k-means++: The Advantages of Careful Seeding. *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 1027-1035.
- Babayan, P.V., Skonnikov, P.N., Trofimov, D.V., 2020: Algorithm of Object Segmentation in the Image for Reference Size Definition While Automatic Object Tracking. *Proceedings of the International scientific forum "Modern Technologies in Science and Education"*, 184-187.
- Barron, J.L., Fleet, D.J., Beauchemin, S.S., Burkitt, T.A., 1992: Performance of Optical Flow Techniques. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 236-242.
- Delgado, M.F., Nguyen, N.T., Cox, T.A., Singh, K., Lee, D.A., Dueker, D.K., ... Samples, J.R., 2002: Automated perimetry. *Ophthalmology*, 109(12), 2362-2374. doi.org/10.1016/S0161-6420(02)01726-8.
- Dutta, A., Mondal, A., Dey, N., Sen, S., Moraru, L., Hassanien, A.E., 2019: Vision Tracking: A Survey of the State-of-the-Art. *SN Computer Science*, 1(57), 1-19. doi.org/10.1007/s42979-019-0059-z.
- Ester, M., Kriegel, H.P., Sander, J., Xiaowei, X., 1996: A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery in Databases and Data Mining*, 226-231.
- Farneback, G., 2003: Two-frame motion estimation based on polynomial expansion. *Scandinavian conference on Image analysis*, 363-370.
- Fischler, M.A., Bolles, R.C., 1981: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381-395. doi.org/10.1145/358669.358692.
- Gonzalez, R.C., Woods, R. E., 2008: *Digital Image Processing: International Version 3rd Edition*. Pearson Education.
- Gonzalez, R.C., Woods, R.E., Eddins, S.L., 2020: *Digital Image Processing Using MATLAB: 3rd edition*. Gatesmark Publishing.
- Otsu, N., 1979: A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9, 1. 62-66. doi.org/10.1109/TSMC.1979.4310076.
- Sun, D., Roth, S., Black, M.J., 2010: Secrets of Optical Flow Estimation and Their Principles. *IEEE computer society conference on computer vision and pattern recognition*. 2432-2439.
- Van Assen, H.C., Egmont-Petersen, M., Reiber, J.H.C., 2002: Accurate object localization in gray level images using the center of gravity measure: accuracy versus precision. *IEEE Transactions on Image Processing*, 11(12), 1379-1384. doi.org/10.1109/TIP.2002.806250.
- Yilmaz, A., Javed, O., Shah, M., 2006: Object Tracking: A Survey. *ACM Comput. Surv.* 38(4), Article 13. doi.org/10.1145/1177352.1177355.
- Zaitoun, N.M., Aqel, M.J., 2015: Survey on Image Segmentation Techniques. *Procedia Computer Science*, 65, 797-806. doi.org/10.1016/j.procs.2015.09.027.
- Zhao, Y., Shi, H., Chen, X., Li, X., Wang, C., 2015: An overview of object detection and tracking. *2015 IEEE International Conference on Information and Automation*. 280-286. doi.org/10.1109/ICInfA.2015.7279299