Detection and Tracking of the Objects in Real Time Applied to the Problem of the Log Volume Estimation

Artem V. Kruglov[1], Yurii V. Chiryshev[1]

[1]Ural Federal University named after the first President of Russia B. N. Yeltsin, Mira st., 19, Yekaterinburg

avkruglov@yandex.ru, iurii.chiryshev@mail.ru

Abstract The paper presents a method for solving the detection and tracking of target objects in real-time machine vision systems in the problem of measuring the volume of logs moving on the conveyor belt. The mathematical processing steps which are discussed in this paper include: 1) Isolation of the foreground objects. 2) Classification and tracking of moving objects. An essential feature of the realization of this task is optimization of the data processing methods by parallelizing them on the GPU. Developed algorithm was tested in the laboratory and the obtained results allow counting on the high accuracy of the detection and tracking module in the real-time system for calculating the parameters of roundwood. All components of the module are designed using common methods to achieve the high-speed processing of visual data among operating systems MS Windows.

Keywords: Machine vision, image processing, background model, parallel processing, real-time system, log detection

1   Introduction

The basis of the most programmable moving detectors is the method of calculating of the interframe differences and constructing a binary mask of motion for each frame followed by a search for moving objects on it (Lukjanica, Shishkin, 2009; Forsyth, Ponce, 2012). More intelligent methods include, for example, algorithms for computing foreground objects as well as their shadows or algorithms, which adapt to the background dynamically changing (Zivkovic, 2004, 2006; Prati, Mikic, 2003). Since the projectable measurement system can be used in extremely harsh conditions of the industrial enterprise with a priori uncertain external influences (e.g., movement of the conveyor elements), for solving the problem of precise localization of moving logs the method of the separation of foreground and background based on a probabilistic model was developed.

2 Background model creation and selection of the foreground objects

In the proposed method, the pixel values changing over time is considered as a pixel process, i.e. time series. In this case, changes of the background pixels are modelled using multivariate normal distribution with density:

$$\eta(\mu_t, \Sigma_t) = \frac{1}{(2\pi)^{n/2}(\Sigma)^{1/2}} exp\left(-\frac{1}{2}(X_t - \mu_t)^T \Sigma^{-1}(X_t - \mu_t)\right) \qquad (1)$$

where μ_t – expected mean, Σ_t - co-variance of t component and χ_t - pixel value.

The process of forming the background model and estimating its parameters μ_t and Σ_t performed on several consecutive frames, using the assumption of independence of the intensity values through various channels. Nonsingular normal distribution is calculated for each background pixel.

$$P(X_t) = \sum_{i=1}^{K} \omega_{i,t} \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \qquad (2)$$

Each term in the sum corresponds to the process in the pixel stage, which is characterized by the parameters of the normal distribution with a weighting factor ω_i, which is a frequency coming of this process in a given pixel to the attention of the camera. Parameter K is the maximum number of distributions of the mix (in this problem K is equal to 5).

In order to prevent the segmentation of shady areas to the foreground the method (Prati, Mikic, 2003) was used. This method is based on the hypothesis that the most important hallmark of the shadow image, distinguishing it from the image of the real scene is transparency, i.e., the background features are clearly visible through the shadow which is applied to them, although the overall illumination of this part of the scene is significantly reduced. Therefore, in order to separate the object and shadow, it is necessary to check each pixel of the foreground for transparency, for which purpose the procedure of comparing pixels of the original image and the background model with taking into account changes in illumination is performed. Figure 1 shows the results of the algorithm implementation to the segmentation of the video into three classes: foreground, background and shadow.
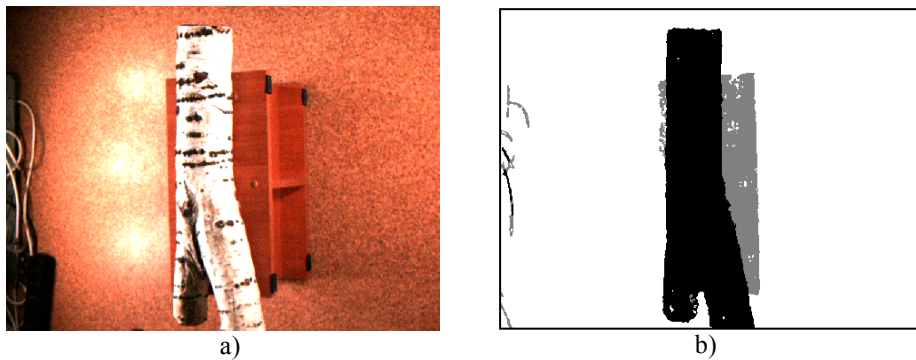

a)   b)

Fig. 1. Result of foreground objects selection a) video frame; b) foreground mask (black colour corresponds to foreground object, grey colour – shadow and white – background)

To periodically update of the background model the following strategy is used. A new pixel value is added to the background sequence only if it was segmented as a background in the previous frame. This mechanism increases the probability of correct determination of the background in subsequent times for the prevention of classifying log's pixels to the background or shadows when the log stops under the camera.

Thus the proposed algorithm performs the following technology for separation background and foreground:

- algorithm is able to detect the movement of shadows;
- the ability to automatically adjust of algorithm to the periodic motions existing on the video after a certain period of time;
- algorithm is able to reflect the slow changes of illumination.

3   Detection and tracking of the logs

Logs detection can be considered as a special case of recognition problem which has only two classes of objects – "log" and "non log". Many successful methods of solving such problems (recognition) with a variety of practical applications have been created over the past decade. However, in most cases, they are applicable for certain special cases, such as tracking of faces, vehicles or license plates and cannot be used for this task. Therefore, for the problem of detection and isolation of logs the specific methods were developed.

The problem of detection and localization of logs in the video image can be formulated as follows: in the frame of the video sequence the presence or absence of logs is ascertained and in the affirmative case the boundaries of the rectangular frame comprising the entire log are found. The solution to this problem should be invariant to the changeable conditions of the scene around the log (for example, small changes in lighting, movement of foreign objects or adjacent logs in frame).

Note that the construction of the system that meets all these requirements is an extremely difficult task. In addition, the selection of logs in the video sequence is complicated by the following factors:

- logs detection should be carried out in real time, which imposes much tighter restrictions than the methods of selecting objects on the static images;
- one log is observed in a series of consecutive images, so it is necessary to track the logs between frames in conjunction with the detection;
- number of logs in the image generally is not known in advance ;

Based on analysis of a large database of video images of real technological process the logs detector based on the implementation of background patterns and several statistical methods (Lukjanica, Shishkin, 2009; Zivkovic, 2004) has been developed.

With preliminary foreground image obtained in the previous step (based on the model of the background), you can use the simple, but fast methods of morphological operations to remove noise. Then the connected components are combined into the blobs and the smallest possible bounding rectangle is found (Wu, Otoo, Suzuki, 2008), herewith too small and shadow areas of an image are not taking into account.

After the foreground objects are allocated, they have to be compared with the objects from the previous frame. It is performed on the basis of similarity of shape and location (computed by finding the intersection of the bounding rectangles of two ob-

jects) In general when the correspondence between the objects of successive frames is established algorithm fulfils several different cases:

- definite correspondence between objects was found
- correspondence to the several objects of the previous frame was found (such case is possible when two foreground objects overlap)
- no correspondence was found

Each detected object has two attributes: age and identity. The former is the number of frames in which the object was observed, and the latter is the serial number of the detected object (eventually this number will be the identifier of the log).

To simulate the behaviour of moving logs, namely track events: the entrance to the measuring zone, being in the measurement zone and exit from the measurement zone Bayesian approach was used in which these events were considered as hypothesis (Lukjanuca, Shishkin, 2009).

Measuring zone - an area within an image, usually located in the centre of the frame with smaller than the whole image size (see Fig. 2). The necessity of implementation of the measurement zone is in the following reasons: a log - is an elongated object which is observed on a number of consecutive frames depending on its speed, and most of the time the log is located in the centre of the image and available for the measurement of its geometrical parameters. However, there are periods of time when the log is not suitable for measurement due to the small area of its silhouette even it is observed in the frame. As a rule, these periods relate to the time of entry and exit of the log from the frame. At such moments the log is tracked, but the no measuring is performed.

There are four possible hypotheses in terms of location of the object for a moving object, and a measuring zone :

- H1: log outside of the zone;
- H2: log entering the zone;
- H3: log inside of the zone;
- H4: log leaving the zone.

Herewith consider two criteria:

- distance between the object and the zone a few frames back;
- distance between the object and the zone at the moment.

Further procedure of logs identification uses the results of image segmentation as input data which includes a number of found objects, their position and size (bounding rectangle). At first all pairs object-zone with detected event "zone entering" are considered for the correct identification of the log. Such objects are considered further in terms of their compliance with the event "zone leaving".

Thus, the object that observed in the sequence of frames considered as a log passed through the measuring system if it is associated with the following sequence of events:

1. The object has entered the zone from top/bottom.
2. The object has been being inside the zone for a long time.
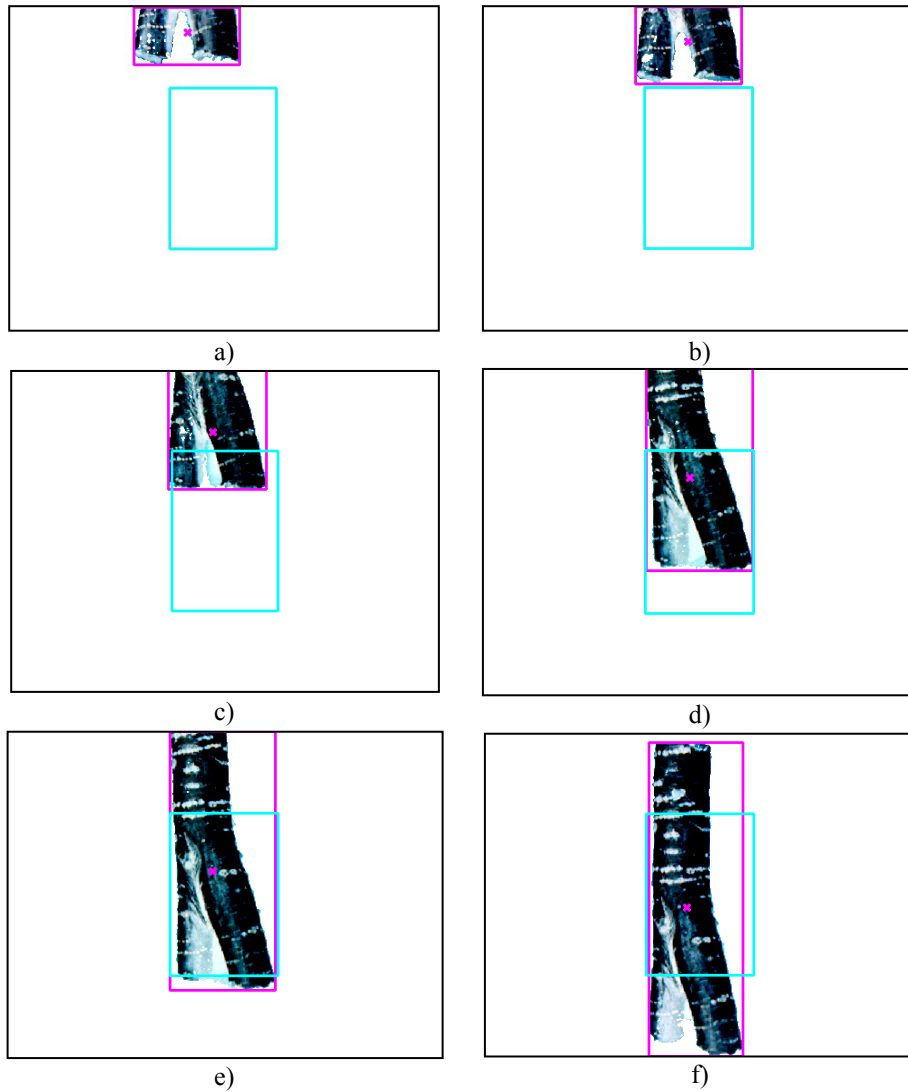3. The object has left out the zone from the bottom/top.

Fig. 2. Video sequence corresponding to the event of entering log (its area bounding by purple rectangle) to the measuring zone (blue rectangle). a) – c) log outside the zone d) – f) log inside the zone.

The final stage of the processing of this module is a matching of the visible parts of logs, observed from two synchronized cameras. In other words, it is necessary to establish for each log the one-to-one correspondence of its image obtained from the first camera to the equal image received from the second camera on the other side of the conveyor. This operation caused by the need of grouping of the available information on the measurement objects in pairs for further stages of feature descriptions formation and measurements of geometric characteristics of the logs.

If the problem of combining pairs of frames is being formulated as an optimization problem it can be narrowed down the one of the fundamental task in the field of mathematical optimization - assignment problem, and solved with the use of the apparatus of combinatorial optimization (Rublee, Rabaud, Konolige, Bradski, 2011) . common the problem is formulated as follows:

Let us consider two sets U and V of the same size, and the cost function C. It is necessary to associate each element of the set U to the exactly one element of the other set $f:U \rightarrow V$, so that the objective function

$$\sum C(u, f(u))$$ (3)

has minimal value.

In the context of this problem it was required that the sum of the Euclidean distances between the pair of logs from two synchronized frames is minimized. Then by defining the data structure in terms of a bipartite graph, the result of the algorithm is a list of graph edges directed from U to V and forming the minimum cost matching. The result of the algorithm of combining the two parts of the log is shown in Figure 3.
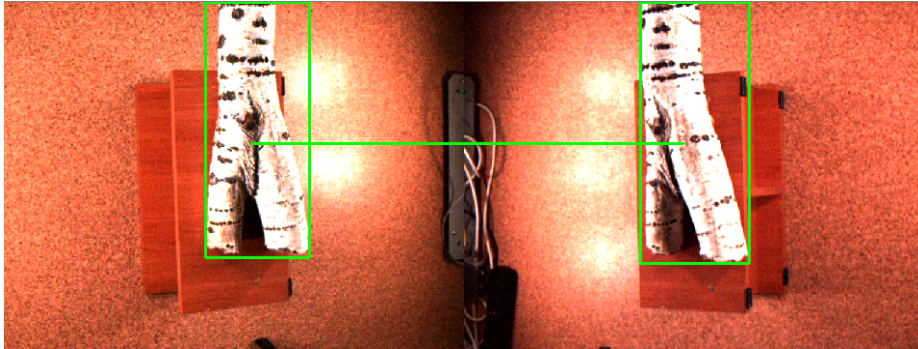

Fig. 3. Visualization of the stereo frames combining

4 Feature descriptions construction and measuring of the geometry

The measuring of geometric characteristics of a log consists of 3 stages:

1. Calculation of diameter and cross-section area along the whole length of a log
2. Calculation of the length of a log
3. Calculation of the net volume of a log

To solve the first problem it was decided to approximate the log with a generalized cylinder and use the algorithm of the rotation body reconstruction over the two images. It consists of two main steps: firstly, an axis of symmetry of the rotation body in three-dimensional space is constructed over the two images via triangulation; secondly, the cross sections of the reconstructed model of the rotation body are computed iteratively, and with help of these cross sections the model of the rotation body is constructed.

During the preparation for the reconstruction it is necessary to underline the boundaries of the rotation body in the input images. It is assumed that the segmentation was successful and marked the boundaries of the object completely. Further, the boundaries are described by smooth curves, which will be used to build the axis of symmetry. An algorithm for constructing the axis of symmetry of the rotation body is based on the assumption that the axis of symmetry in three-dimensional space is projected on the axis of symmetry of the rotational body on each of the input images (Fig. 4). In fact, the problem is reduced to the definition of the 3D coordinates through the known 2D coordinates of the correlated curve points. At this the epipolar constraint imposes on these points (Wöhler, 2009; Bradski, Kaehler, 2008).
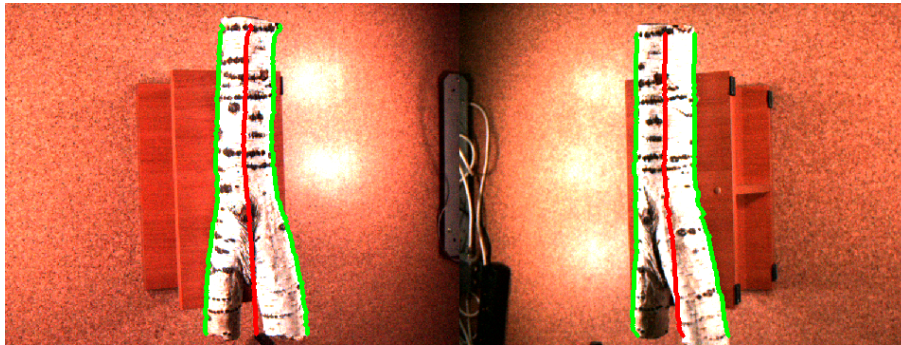

Fig. 4. The boundaries and the axis of symmetry of the log

Solution of the problem of calculating the length of the logs based on the assumption that for a video containing an object in motion, it is possible to calculate the direction and velocity in several points of the log, i.e., calculate the optical flow (Lukjanica, Shishkin, 2009; Forsyth, Ponce, 2012). At this the object of measurement has the following restrictions: log like a solid body moving in the plane of the conveyor has a maximum of four degrees of freedom (two rotational and two translational). In this case the length of a log will be the integral sum of shifts measured at discrete times during the passage of the log through the measuring system.

For the calculation of the correspondences in the sequence of frames received during the movement of the logs through the measuring system the methods based on searching and comparing the tie points were applied. The general scheme of image matching techniques can be described by the following sequence of actions

- detecting characteristic points on the image;
- formation of descriptors in the neighbourhood of characteristic points;
- comparison of descriptors based on the specific metrics;
- filtering matches based on the specific prior model.

Thus the first two stages of image matching is performed per-pixel image processing. At the final stage the comparison of pairs of images and exclusion of the improperly founded correspondences according to the specific constraint model are implemented.

Assumed the specificity of the task the applied matching algorithm is the following:

1. For the detection of low-level features in the images two detectors of the corner points have been considered - FAST and Harris detector (Rosten, Porter, Drummond, 2009; Rosten, Drummond, 2006; Harris, Stephens, 1988). Due to the better resistance to noise, the Harris corner point detector is applied. It based on the searching the regions where the local curvature in the sample window has its maximum (Fig. 5).
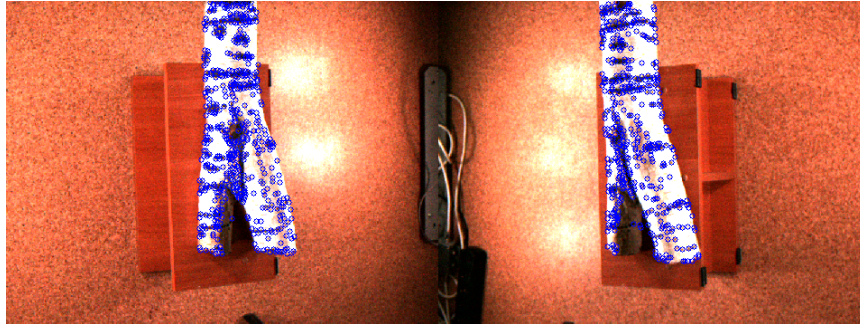

Fig. 5. An example of the frames processed with Harris detector. Blue circles show singular points

2. The window-based locally binary descriptor ORB (Rublee, Rabaud, Konolige, Bradski, 2011) was choose as a descriptor of a characteristic point. An essential feature of this descriptor is its invariance to rotation and change in brightness of the characteristic points. As a metric for comparing descriptors the Hamming distance is used.

3. In order to ensure sustained and rapid points matching the pyramidal (hierarchical) optical flow algorithm is used. The result of the algorithm is an array of vectors, whose elements correspond to the direction and movement of each characteristic point (Fig.6). The method of the sustainable estimating of the model parameters RANSAC (Forsyth, Ponce, 2012; Wöhler, 2009) is implemented to filter the motion vectors from the noise. The method is based on the criterion of the best matching over the whole set of pairs with the selected model which parameters are estimated by a random sampling from the entire set of pairs. Thus, the quality of matches between features of different images depending on the model of physical limitations.

To solve the problem of the stable assessment of matching sets of the points pairs two models were considered:

- fundamental matrix, which used under the supervision of the general scene;
- homography, which can be applied on the assumption that the observed scene can be approximated by a plane.

The general scene is used as a geometric model of transformation via RANSAC scheme for adjacent frames. Those the fundamental matrix which most closely match the observed motion from the view of epipolar constraints (Wöhler, 2009; Hartley, Zisserman, 2003) was adjusted during the iterative refinement.

Repeating of this procedure for the each frame of a video sequence allows determine (accumulate) the full length of the log L, as the integral sum of the separate displacements recorded during the passage of the log through the measuring system.
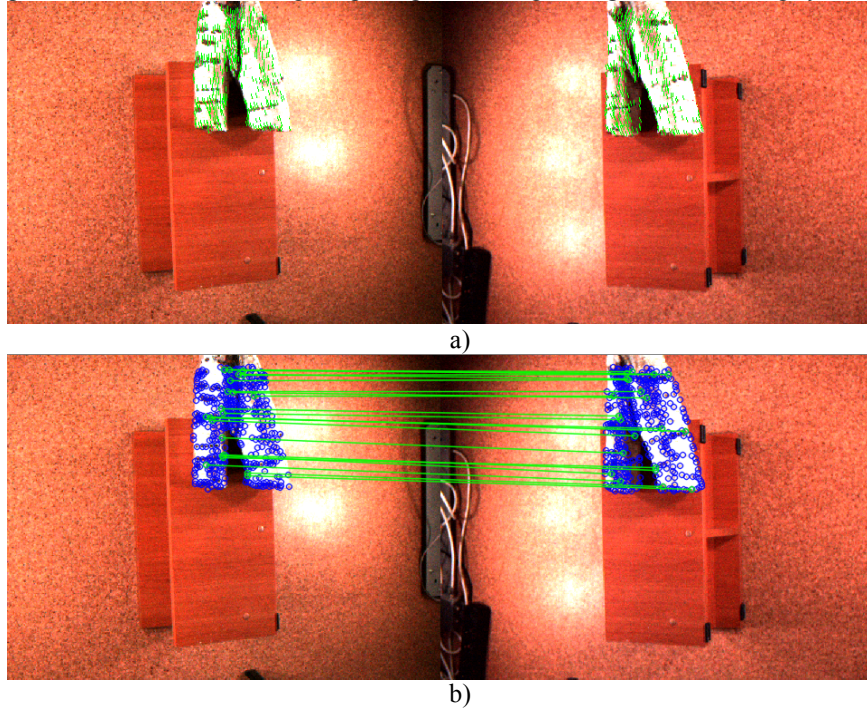

a)

b)

Fig. 6. Example of the correlation of the characteristic points a) combining consecutive frames (optical flow) via RANSAC scheme; b) matching the right and left frame.

Approximating the measuring object by a set of cylinders with the unit height i, the net volume of the log will be calculated as the integral sum of the cross-sectional areas via the well-known equation for the volume of the rotation body:

$$V = \pi \sum_{i=0}^{L} r_i^2 \qquad (4)$$

where r_i - radius which forms a body of revolution.

5   Conclusion

In conclusion, the offered module including algorithms for determining the geometric parameters of logs was programmed in MS Visual C ++. The module was tested on the real video sequences and high efficiency in detection of the moving logs without tracking failures. The performance of the module allows to process a video sequence with a frequency of 20 frames per second (tests were implemented on the PC Intel Core i7, 2800 Mhz), which makes possible to use the solution to determine timber volume in the real-time systems.

References

1. A. Lukjanica, A. Shishkin, Digital Image Processing. I-S-S Press, Russia (2009)
2. David A. Forsyth, Jean Ponce Computer Vision: A Modern Approach, 2nd edn. Pearson (2012)
3. Kesheng Wu, Ekow Otoo, Kenji Suzuki, 2008. Two Strategies to Speed up Connected Component Labeling Algorithms. LBNL Tech Report, LBNL –59102
4. Z.Zivkovic Improved adaptive Gausian mixture model for background subtraction International Conference Pattern Recognition, UK, August, 2004
5. Z.Zivkovic, F. van der Heijden Image Pixel for the Task of Background Subtraction Pattern Recognition Letters, vol. 27, no. 7, pages 773-780, 2006.
6. Prati, A., Mikic, C., Trivedi, M. M., Cucchiara, R. (2003) Detecting Moving Shadows: Algorithms and Evaluation. In: IEEE Transactions on Pattern Analysis and Machine Intelligence. vol. 25. no. 7. pp. 918 923
7. Wagner H. M., Principles of Operations Research. Prentice-Hall, Englewood Cliffs (1969).
8. C. Wöhler. 3D Computer Vision: Efficient Methods and Applications., 2009
9. Machine learning for high-speed corner detection, E. Rosten and T. Drummond, ECCV 2006
10. Hartley, R. I. and Zisserman, A. Multiple View Geometry in Computer Vision, Cambridge University Press, 2003
11. Gary Bradski, Adrian Kaehler. Learning OpenCV, O'Reilly Media, 2008
12. Senior, A. Hampapur, Y-L Tian, L. Brown, S. Pankanti, R. Bolle. Appearance Models for Occlusion Handling. Second International workshop on Performance Evaluation of Tracking and Surveillance Systems & CVPR'01. December, 2001.
13. C. Harris, M. Stephens. A combined corner and edge detector. Proceedings of the 4th Alvey Vision Conference: pages 147—151. 1988.
14. E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. Orb: an efficient alternative to sift or surf. 2011.