2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS 2019)

Object Detection and Tracking Approaches for Video Surveillance Over Camera Network

1st Nitesh Funde
Department of Computer Science
Visvesvaraya National Inst. of Technology
Nagpur, Maharashtra, India
nitesh.funde@students.vnit.ac.in

2nd Parnika Paranjape
Department of Computer Science
Visvesvaraya National Inst. of Technology
Nagpur, Maharashtra, India

3rd Kamal Ram
Amazon Development Centre India
Hyderabad, Telangana, India

4th Punit Magde
FactSet Systems Pvt Ltd.
Hyderabad, Telangana, India

5th Meera Dhabu
Department of Computer Science
Visvesvaraya National Inst. of Technology
Nagpur, Maharashtra, India

***Abstract*—Object detection and tracking are the most challenging part of any computer vision applications. In computer vision, video surveillance is a popular research area in a dynamic environment, particularly for security reasons. The video surveillance technology plays a crucial role to prevent crime, terrorism etc. The video outputs are filtered and processed by human operators and in case of a forensic, the high volume of data made it difficult to track any object. This work has been done with an aim to reduce the effort of human operators with an increase in the response time to forensic events. It involves designing of an efficient object tracking system for simple environments where the camera is static, background is simple and no similar object to the one being tracked is present. The system is provided with network configuration of the cameras and roads of the surveillance area, videodumps and an image of the object to be tracked. It tracks the objects through the videos and dumps the tracked portions of the videos where the object was present.**

***Index Terms*—Object Detection; Object Tracking; Image Segmentation; Video Frame; Video Surveillance.**

I. Introduction

In computer vision, video surveillance is a challenging research area which detect, discover, recognize and track objects in videos. Videos are defined as sequences of images where each image is a frame which is showed with its continued fast frequency such that human eyes can view the content in continuous way. In general, the content of two subsequent frames are closely related.

The object detection is about finding object location in videos. It includes the verification of object presence in sequences of images and locating object accurately for recognition. Object tracking involves monitoring an objects path in a sequence of images with details such as size, position, shape etc. The process of object detection is performed by matching target areas in a sequence of frames which are at closely related time periods. These processes i.e. object detection and tracking are related closely as object tracking initiates with detection of objects whereas object detection repeatedly in subsequent frame sequence is required to verify object tracking. Thus, every object tracking needs an object detection methods in each and every frame.

The availability of higher quality video cameras and the high performance computers are required for the automated video analysis in computer vision applications. There are several applications in residential, industrial, commercial categories such as surveillance, urban planning, security and traffic control etc where object detection and tracking are our utmost interest. These applications still need to be studied for its role in consumer electronics. They need complicated installations, good quality of requirements, setup procedures and an expensive and specialized hardware for working satisfactory conditions. In general, a video surveillance system consist of video cameras fixed to a position or pan-tilt and transmit video data stream to a control room where human operators are monitoring and tracing video data streams. However, it is always difficult for human operators to detect an object which is moving in the feasible region of cameras where it is the possibility that the human operator may miss the object. Therefore, monitoring video data streams is a challenging task when a number of cameras in such a video surveillance system network increases. It is necessary to automate the video surveillance systems to eliminate such human operator errors and improve the performance.

978-1-5386-9533-3/19/$31.00 ©2019 IEEE

A. Objective

Automated object tracking are essential for several important applications. The higher level computer vision based intelligent applications need a perfect and efficient tracking capability in a video surveillance system. The objective of video tracking is to associate target object to a sequence of frames in videos. In general, the building relation is not easier when objects are moving with fast rate as compared to the video frame rate.

The main objective of this paper is to build up a standalone(offline) application with the following characteristics:

1) The application provides a support for taking the geometry of the network(roads and cameras) of the surveillance area as input.
2) Videos from all the cameras are also given as an input.
3) User can play any video and select an object from that by pausing and cropping his region of interest as the target object need be tracked in the provided videos .
4) The application automatically detects the next video to be played when the target goes out of frame in the current video.
5) Finally, the application outputs the segments of the videos where the target was tracked with time-stamp on it.

II. Object Detection Methods

The first stage of object tracking is to track objects which are of our utmost interest in a sequence of frames and clustering the pixels of these objects . As we know moving object are somewhat difficult to track, therefore; number of techniques have emphasized on moving object detection. There are several methods described below.

A. Frame Differencing

Videos are defined as sequences of images in which each image is called a frame. We can use frame differencing technique for moving object detection in video surveillance. The differentiation between the frame at current time and reference base frame is known as a background frame. This method is known as frame differencing method [1]. If the difference resulted into changing pixels, then it may indicate that the object is moving. The object moving detection is performed by checking the difference between two successive frames. We can adapt this method in dynamic environment. In general, it is not easy to obtain a total path of a moving object [2].

The technique of frame differencing include sequence of three frames which is in between frames F_{k+1} and F_{k-1} [3]. The difference between these frames are calculated as d_{k+1} and d_{k+1} expressed as below:

d_{k+1} = $|F_{k+1} - F_k|$

d_{k-1} = $|F_k - F_{k-1}|$

Now these values are converted to boolean values by comparing them with threshold T.

$$D_{k'}(x,y) = \begin{cases} 1, & \text{if } d_{k'}(x,y) \geq T \\ 0, & \text{otherwise} \end{cases}$$

After this procedure, binary AND operation on D_{k+1} and D_{k-1} is performed.

B. Optical Flow

The technique of optical flow [4] is used to calculate field of optical flow of image and to cluster which is depend upon the optical flow characteristics of image distribution. The results are the detailed information of each movement and object moving detection. However, in case real-time dynamic environment, poor anti-noise performance and sensitivity to noise etc. issues need to be handled. A dense displacement vector represents the each and every pixel in an area. It is calculated by using the constraints such as brightness of respective pixels in subsequent frames. In motion segmentation, optical flow is generally utilized in terms of feature.

C. Background Subtraction

It is the renowned technique for motion detection. Here, results are obtained on subtraction of the current image from a base background image which is updating in an every time interval. This technique is performing well in case of fixed cameras. The result will be new objects which consist of complete silhouette area of an object. This technique is simple and sensitive to dynamic environment situations such as lightning events. Thus, this technique is rely on a well background model.

The background modelling is first and basic stage for background subtraction. It should be sensitive for identifying moving objects. The background modelling is considered as a reference model. The reference model is taken in this technique in which each frame is checked for determining any variation. The variations represents the presence of moving objects. For background modelling, median and mean filter [5] are used. In [6], background subtraction have describe as two types i.e. recursive and nonrecursive. For background estimation, recursive techniques [6], [7] is not maintaining a buffer. The sliding-window approach is used by non-recursive technique [6], [7] for background estimation.

There are mainly two background subtraction methods. [3]

1) Simple Background Subtraction: The differentiate between each I_t(x, y) current image and base background image is used for the motion detection D(x, y). In general, the first image of video is the reference image.

$$D(x,y) = \begin{cases} 1, & \text{if } |I_t(x,y) - B(x,y)| \geq \tau \\ 0, & \text{otherwise} \end{cases}$$

Here τ is a threshold which determine the type of pixel i.e. background or foreground. The foreground pixel is present when difference is greater than or equal to τ otherwise it is categorized as background pixel.

2) Running Average: The variation in illumination resulting in noise in simple background subtraction method. Thus, the issue of noise can be solved if the background is adaptive to changes in temporal view which is updated in each frame

$$B_t(x,y) = (1-\alpha)B_{t-1}(x,y) + \alpha I_t(x,y)$$

where α is a learning rate. The $D(x, y)$ denotes binary motion detection mask as given below:

$$D(x,y) = \begin{cases} 1, & if|I_t(x,y) - B(x,y)| \geq \tau \\ 0, & otherwise \end{cases}$$

III. Object Tracking Methods

The discovery of path of a moving object in image plane can be called as tracking. The objective is discovering route of an object by detecting its place in every frame of video. Object tracking is categorized as point tracking, kernel dependent tracking and silhouette based tracking.

A. Point Tracking

The moving objects in a image structure are constituted by various feature points in tracking. The point tracking is a crucial problem particularly in false object detection and in events of occlusions. The recognition is simple relatively using thresholding based approach for the point identification.

1) Kalman Filter: It is depend on the optimal data processing recursive algorithm. It carries out density propagation using restrictive probability. It represents the set of equations which gives computational efficient means for estimating the various states of a process in various perspectives [8]. It provides present, past and future state estimations and even in case of unknown modelled system. This technique calculates the state of process and then take feedback in terms of noisy measurements.
2) Particle Filtering: This method creates almost all models for single variable prior to taking the next incoming variable [9]. It is beneficial when there are large number of variables in a dynamic environment. The resampling can be done for new operation. In kalman filter, state variable are assumed to be of normal distribution. Thus, this technique gives state variables which are of poor approximations. This restriction is solved by particle filtering. The particle filtering uses color features, texture mapping and contours. It is a Bayesian sequential sample methods where the recursive methods for distribution by finite set of weighted features are performed. It has two steps; first is prediction ; second is update same as this technique.
3) Multiple Hypothesis Tracking (MHT) : In MHT algorithm [9], the large number of frames have to be tried for obtaining better outcomes of tracking.The process is using iteration where it start with a group of previous track. A group of disconnected tracks represents a hypothesis and predicts position of object in successive frame. The measured distance is used for predictions. . MHT have capability of handling occlusions, tracking multiple objects and obtaining optimal solutions.

B. Kernel-based Tracking

It is generally executed by calculating embryonic object moving region from one particular frame to the next incoming frame. This is actually done in terms of object motion of parametric type such as affine, conformal and translation. The various algorithm differs in form of representation, selection of method for approximating object motion and count of tracked objects. The geometric shape is utilized for illustrating an in dynamic environment. However, it may happen that some parts of object can be outside of determined shape whereas some part may present inside. The non-rigid and rigid objects are used to detect this situation. The shape, appearance, representation of objects are used for tracking techniques.

1) Simple Template Matching: This technique is same as brute force approach for determining the interest region in the video [10]. Here, a base image is checked with frame which is disconnected from the video. Tracking is performed for a single object in video and it can overlap with the object partially. This technique is used to process digital images to discover small region of interest which matches to a template image. The procedure of matching consist of template image for all feasible solutions in source image and compute an index which specify how the model fits that picture position.
2) Mean Shift Method: This technique is used to discover the region of frame which is identical to a previous model. The histogram represents that region. The procedure of a gradient ascent is essential to shift tracker to a position where it maximizes a score of similarity between the current image region and model. In this technique, representation of target can be of elliptical or rectangular region. The target model is represented by the probability density function. The spatial masking and asymmetric kernel regularized the target model.
3) Support Vector Machine: SVM [11] is a popular classification technique which provides a set of negative and positive training data values. The object image tracked is included in the positive samples whereas all remaining thing not tracked all included in negative sample. With proper training and physical initialization, the partial occlusion of an object can be handled.
4) Layering Based Tracking : Here, in this technique, number of objects are tracked using kernel tracking. Each layer includes representation of shape, appearance , motion such as rotation and translation using intensity. The layers can be formed by compensating the motion background so that motion of an object can be calculated from image rewarded using motion which is 2D parametric. The pixels probability of each objects motion and shape [11].

C. Silhouette Based Tracking Approach

The complicated shape such as fingers, hand and shoulders may present in some objects which can not be defined by simple and general geometric shapes. This method is used for describing the accurate shape of objects. This technique is used to find the region of each object in every frame using an object model represented by previous frames. It has the capability to deal with different type of objet shape, object split and occlusion.

1) Contour Tracking: The technique is progressively iterated a primary step of contour in the existing frame to its new location of current frame. It is necessary that some objects which are present in current frame which is overlay with region where object present in existing frame. There are two different types for contour tracking. In first type, the state space models are used for modelling contour motion and shape. In second type, the direct minimization methods are used for evolving the contour by minimizing contour energy. It is very advantageous to use silhouettes tracking for handling a different types of object shapes.
2) Shape Tracking: The shape tracking is used for investigating an object model present in the previous frame. The performance is identical to template tracking. There is another view to shape matching; it is used to discover silhouette based matching in two continuous frames [9]. It is similar to matching point. The background subtraction perform the silhouette based detection. It has capability to deal with occlusion handling and single object and which is carried out with help Hough transform methods.

IV. Segmentation

The segmentation is defined as a separation of objects from the background. The partitioning of images to similar regions is the main objective of image segmentation. There are two problems ; first is to achieve an efficient partition and second is to choose measures for testing a partition which must be addressed by every segmentation algorithm. The foreground segmentation is about the division of a scene into two classes i.e. background and foreground. The region such as buildings, roads and furniture are considered as background. The appearance is expected to change over the time due to various factors such as weather, lighting conditions whereas the background is fixed. The foreground are the elements which are expected to move and some foreground elements may be stationary for long duration such as parked vehicles.

The different segmentation techniques are discussed below:

A. Mean Shift Clustering

For image segmentation problem, the mean shift clustering is proposed in [12] where it used to discover clusters in spatial and color space jointly $[l, u, v, x, y]$, where $[x, y]$ denotes the spatial location and $[l, u, v]$ denotes the color. There are large number of counts of hypothesized cluster centroids selected randomly for the problem. Each cluster is then shifted to centroid data which may be located in multidimensional ellipsoid centroid on cluster. This process defines a vector which we called as mean-shift vector. This vector is calculated iteratively when we found that cluster centers are not changing their positions. It may happen that some of clusters may merge during iterations.

B. Graph-Cuts

We can consider the problem of image segmentation as a graph-cuts where vertices represents pixel are divided into N separate subgraphs by trimming the edges (weighted) of a graph. Weight is generally calculated by brightness, color and texture similarity among the nodes. A cut is defined as a total weight of edges trimmed between two subgraphs. In [13], the normalized cut is used to solve over segmentation problem. Here, cut is rely on the sum of weight edges. Also, it depends on the fraction of total weights to nodes which are connected in every partition to all nodes.

C. Active Contours

The evolution of a contour which is closed to boundary of objects can make object segmentation. Here, contour is enclosed by object region tightly.

V. Experimental Results

A. Template Matching

It is a technique for searching and discovering the position of a template in a larger image. The function $cv2.matchTemplate()$ in OpenCV [14], is used for template matching. The template image is slided over the input image to compare template with patch of input image. There are different comparisons available in OpenCV. The output is in terms of grayscale image in which how much of similarity neighborhood with the template.

If template image of size (wxh) and input image of size (WxH) then image output is of size (W-w+1, H-h+1). After this, the function $cv2.minMaxLoc()$ [14] is used to discover the position. It can be considered as the upper left corner of rectangle where w is width whereas h is height of rectangle. We can consider it as a region of template. Please refer to Fig: 1 to see that there are many False Positives at $\tau = 80\%$, but at $\tau = 90\%$, the number of False Positives is zero.

B. Nodes Configuration

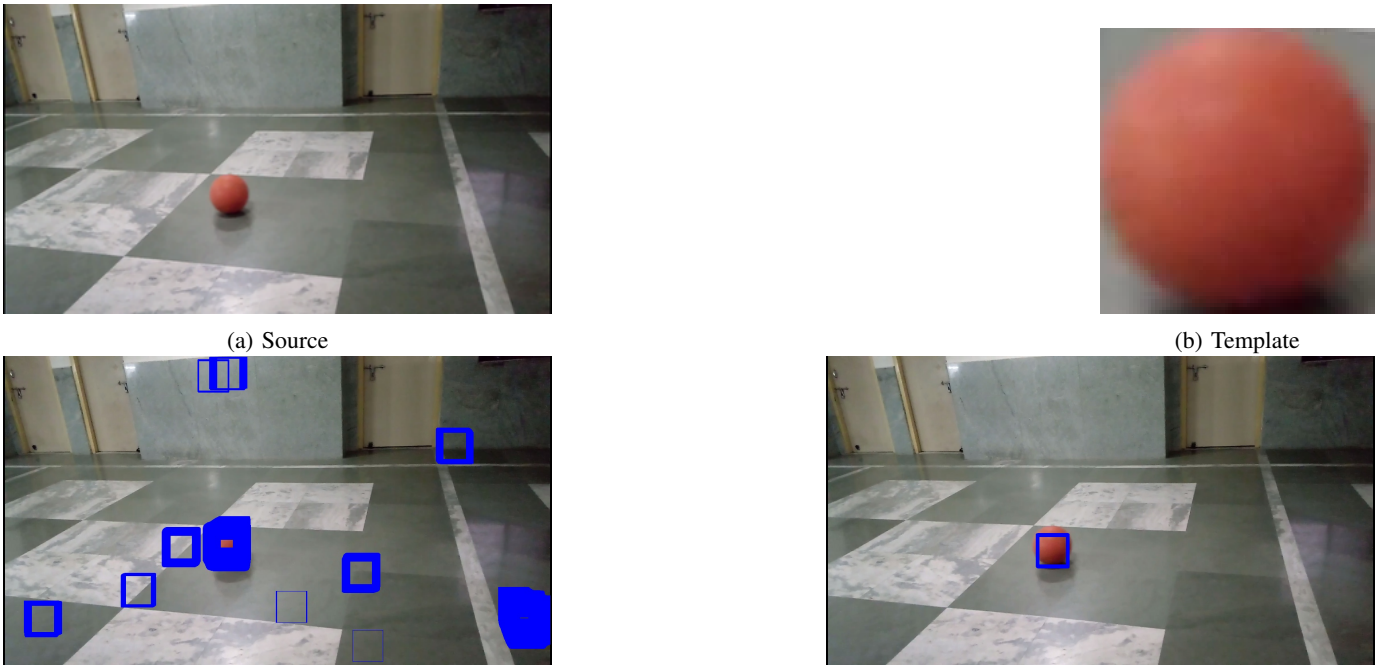
In this section, we take the geometry of the network of area under surveillance as input from the user. The user clicks on the config button on the home screen, and a text-box appears, where he has to input the details in the following format:

- First line is an integer that indicates the number of nodes/cameras. Let it be N.
- The second line contains N integers, where i^{th} integer indicates the degree of i^{th} Node, i.e. the number of roads going out from the Node where i^{th} camera is fixed.
- Let the sum of the N integers in the second line be E, to indicate the number of edges.
- The input is then followed by $2 * E$ lines, each line of which contains 3 integers i.e. a, b and c which indicates that there is a road from Node a to Node b at angle c with respect to the Camera at Node a.

C. Video Selection

Having performed nodes configuration, the user clicks on the Select Source button to choose a source video.

This is the video in which the user will specify a target object to be tracked.



(a) Source (b) Template

(c) $\tau = 80\%$ (d) $\tau = 95\%$

Fig. 1: Template Matching

D. Object Selection

After selecting the source video, the user clicks on the Play button to play the content of the video. At any point, when the user clicks on the Select Object button, the video gets paused, and user has to select a region by clicking on the screen and moving the mouse with left-button clicked, i.e. crop a region. The application automatically tries to figure out the biggest contour in the region selected by the user and draws a boundary around it. The region inside this boundary will be tracked in the subsequent videos.

E. Background Subtraction

Now we need to build a model to separate the background and foreground. This is done by applying background subtraction algorithm on each frame. After separating the foreground, we search the target object in the foreground image, and mark the location where it is tracked.

F. Detection and Tracking

The object that user selects by marking along it's boundary is searched and detected in every frame. The location of the object is also shown on the screen. The object is tracked in all the frames. Whenever the object goes out of frame, i.e. out of coverage area of the camera, it's location is checked and angle is calculated to select the next video to be played. The angle calculation part is shown in the next part. The screen-shots of the object tracked in different videos are shown below:

G. Direction calculation

Whenever the object goes out of frame, we need to calculate it's angle of exit with respect to the current Node's camera frame, so as to decide which video to play next. The angle calculation is shown with respect to the image shown in Fig:2 (c).

Let's assume that the center of the frame is origin. The horizontal line passing through it is x-axis. Now we draw another line to the centroid of the object's last known location, and calculate the angle formed.

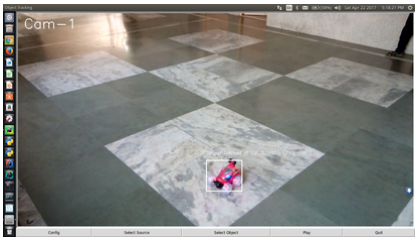
Frame Width = 1312, Frame Height = 703
Center of Frame = Origin = (1312/2, 703/2) = (656, 351.5)
Object Height = 179-104 = 75
Object Width = 59-2 = 57
Centroid of Object = (2+57/2, 104+75/2) = (30.5, 141.5)

Now top-left corner is (0,0), we need to shift our origin.
Thus x = 30.5-656 = -625.5, y = 351.5-141.5 = 210.0
rad = tan^{-1}(y/x), θ = degrees(rad), θ = 161.44

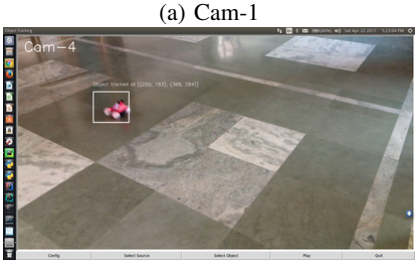
Thus we choose the next video which is approximately located at 160 degree from the camera at the current Node. After all the steps are performed, the application detects and tracks the object in all the frames and stores them as output. Along with the location, the time-stamp is also written on the frames, which shows as to, in what time was the object present at that Node.

VI. Conclusion

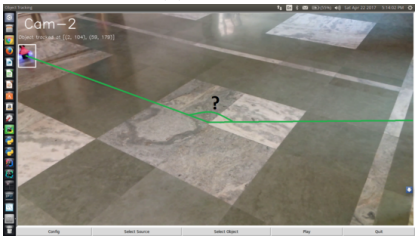
Nowadays, object detection and tracking in videos are studied with the increasing popularity, versatility, availability of video applications. In this paper, the different phases of object detection and tracking have been studied and reviewed. The various methods for these phases are explained in details



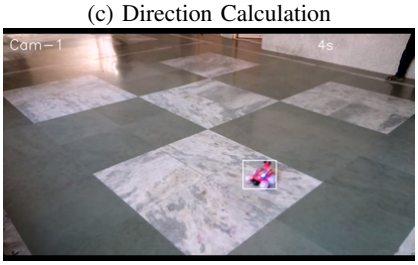
(a) Cam-1



(b) Cam-4



(c) Direction Calculation



(d) Output with time stamp

Fig. 2: Object Detection and Tracking

along with the shortcomings of each and every technique. We developed an application for moving object detection assuming background is simple static and no other similar object to target object is present.

REFERENCES

[1] R. Gupta, "Object detection and tracking in video image," 2014.

[2] R. S. Rakibe and B. D. Patil, "Background subtraction algorithm based human motion detection," *International Journal of scientific and research publications*, vol. 3, no. 5, pp. 2250–3153, 2013.

[3] R. K. Rout, *A survey on object detection and tracking algorithms*. PhD thesis, 2013.

[4] A. K. Chauhan and P. Krishan, "Moving object tracking using gaussian mixture model and optical flow," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 4, 2013.

[5] M. Sankari and C. Meena, "Estimation of dynamic background and object detection in noisy visual surveillance," *International Journal of Advanced Computer Sciences and Applications*, vol. 6, no. 2, 2011.

[6] S. C. Sen-Ching and C. Kamath, "Robust techniques for background subtraction in urban traffic video," in *Electronic Imaging 2004*, pp. 881–892, International Society for Optics and Photonics, 2004.

[7] K. Srinivasan, K. Porkumaran, and G. Sainarayanan, "Improved background subtraction techniques for security in video applications," in *Anti-counterfeiting, Security, and Identification in Communication, 2009. ASID 2009. 3rd International Conference on*, pp. 114–117, IEEE, 2009.

[8] G. Welch and G. Bishop, "An introduction to the kalman filter," 1995.

[9] J. J. Athanesious and P. Suresh, "Systematic survey on object tracking methods in video," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 1, no. 8, pp. 242–247, 2012.

[10] S. Saravanakumar, A. Vadivel, and C. S. Ahmed, "Multiple human object tracking using background subtraction and shadow removal techniques," in *Signal and Image Processing (ICSIP), 2010 International Conference on*, pp. 79–84, IEEE, 2010.

[11] R. Mishra, M. K. Chouhan, and D. D. Nitnawwre, "Multiple object tracking by kernel based centroid method for improve localization," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 137, 2012.

[12] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 25, no. 5, pp. 564–577, 2003.

[13] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[14] O. Team, "Opencv," 2016.