Guillaume Khenchaff's Measure for Clustering Method

Paula Niels Spinoza, Andriamasinoro Rahajaniaina
Dept. of Mathematics, Computer Science and Applications
University of Toamasina
Toamasina, Madagascar
e-mail: nielsspinozapaula@gmail.com,
hajatoam@gmail.com

Jean-Pierre Jessel
IRIT, REVA
Paul Sabatier University
Toulouse, France
e-mail: jessel@irit.fr

***Abstract*—"a contrario" is one of the techniques for tracking objects in real time. However, decomposition methods may still fail to effectively group salient objects according to its movement on the real scene. In this paper, we present an approach for optimizing the "a contrario" grouping method. We introduce a new clustering framework using the probabilistic quality measurement technique, which measures the degree of dependence between the mobile group accepted by the Number of False Alarms (NFA) measure and the group considered to be static in the binary tree. We demonstrate the effectiveness of our method with different situations in uncontrolled environments. We also show its applicability with the Simultaneous Localization And Mapping and Moving Objects Tracking (SLAMMOT) approach.**

Keywords-Salient object; a contrario grouping; probabilistic quality measurement.

I. INTRODUCTION

The successive images processing of a video stream make it possible to incrementally reconstruct a precise 3D model of the scene. In image processing, most of the change detection approaches are based on the interpretation of the difference between a current (or previous) image and a background (image without change or object of interest). Several works focused on updating the background of the image or background subtraction methods [19][20]. However, these approaches are subject to a drift in the estimation of the pose of the moving camera and therefore in the estimation of the movement of salient objects on the scene in a vast environment.

In order to use the "a contrario" algorithm [1] in our approach, we need the points of interest and their information. To do this, we use the Kanade-Lucas-Tomasi (KLT) technique proposed by Lucas et al [13] and then, modified by Shi et al [16]. The analysis of the scattered optical flow behavior obtained from this one module on the captured scene allows us to deduce that the greater the N_{im} images number processed, the better will be the perception of the objects displacements. We use 8 images because these are enough to estimate the apparent movement of a point when using a low cost camera as a test platform. The first two images are used to detect points. Once the points are extracted and tracked by this technique, it is important to distinguish among the tracking points, those corresponding to 3D points attached on mobile objects. The grouping or clustering techniques are required for grouping these ones, but most of them require a priori knowledge of the scene for example the groups number to find as K-means [8] and Mean-shift [6]. The success of these methods strongly depends on these initialization parameters. The same problem occurred in [1] for the analysis of short video sequences. They presented a grouping algorithm based on "a contrario" method, which does not need any parameter or initial information to find in a sequence of images groups of points, which are the projections of 3D points attached on mobile objects.

The work presented in this paper provides an optimization of the "a contrario" grouping technique in order to have relevant information on static objects in an uncontrolled scene. The remainder of this paper is organized as follows. Section II describes an overview of the previews work. Section III presents our contribution; Section IV explains our experiences and our results. The conclusion and future work close the paper.

II. RELATED WORK

Several researchers are trying to solve the problem of automatically finding alignments in a set of 2D points. In this section, we present the related works to the tracking of objects in a real scene. Distinguishing dynamic objects requires knowledge of their speed, orientation and position.
Application of computer vision techniques for autonomous cars is described in [3]. Buyval et al. [4] used a real-time vehicle and pedestrian tracking technique and [18] adopted real-time human object tracking for intelligent monitoring. These approaches are based on an offline tracking technique. However, for a disturbed scene or the camera encounters a difficult situation such as lighting problem, partial or total occlusion, motion blur, etc. it's necessary to make object online tracking [5][12][17]. There are also different techniques that used visual data [4][15]. In general, visual tracking of objects is a problem of computer vision, above all, when the objects or events to be detected are multiple, of variable forms and poorly understood. In fact,

 ISBN: 978-1-61208-779-5

these approaches above need other methods to efficiently group images pixels according to their local texture such as the one discussed in [8]. Gomez at al. [2] proposed a correct alignment detection, which depends on the amount of masking in the texture, the bilateral local density of the alignment, internal regularity and reduction of redundancy. Nebehay et al. [14] described a matching method for deformable objects for single target object tracking.

These different approaches offered advantages such as a minimum amount of background pixels [11], tighter data sets, obtaining an object's orientation in the image plane. However, there are still problems to be solved: computation of rotation angle and scale estimation. Several researches tried to give a solution to these problems, but there are still limits in terms of tracking speed or accuracy [10][15]. The main goal of our work is to solve these ones. We propose an improved of the "a contrario" grouping technique by integrating to it data mining technique called Measure by Guillaume Khenchaff (M_{GK}) [7].

III. CONTRIBUTION

The presence of dynamic objects in an uncontrolled environment could distort the topological map of SLAM. It will be necessary to adopt a grouping technique capable of grouping these mobile objects and optimizing static objects. To solve this problem, we chose the “a contrario” technique.

A. Évaluation of the backgroud model

The objective of the “a contrario” is to group points of interest having a coherent movement along a short sequence of images. Here, the consistency criterion refers to motion vectors which have roughly similar magnitudes and directions for all the points of the group. The method receives a set V of input vectors (x, y, v, Θ | t) where x and y represent the magnitude and v the orientation, which is defined in R^4. The latter contains the scattered optical flow accumulated points of interest over time. In the vector V, the variable t is added just to indicate the moment when these points were selected (start of the tracking time).

The first objective consists in evaluating which elements of V have a particular distribution and contrary to that established by the background model. To avoid element by element evaluation, a binary tree is constructed with the elements of V using the simple link method to have all the groups that can be formed from these elements. Figure 1 shows a graphical representation of a binary tree constructed from 8 points. We find in the root the group, which integrates all the elements of V; on the leaves, the elements where each group contains a single point. Each node in the tree represents a candidate group of points G (x, y, v, Θ | t) ⊂V, which will be compared with the background model using a set of regions pre-established in R^4.

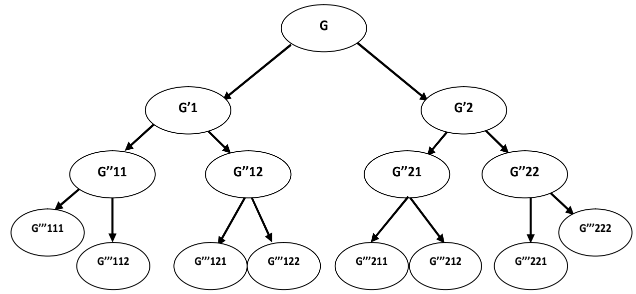


Figure 1. Binary tree of 8 points of interest.

A set of test regions $\mathcal{H}$ must be established in order to evaluate the distribution function of each group G of tracked points resulting from the accumulation of the optical flow where G⊂V. The region space $\mathcal{H}$ is used to calculate the probability that the distribution of each group in the binary tree is similar to the distribution of a model for background objects. In the background model establishes a random organization of the observations distributed in an identical and independent manner and which follow a p distribution.

For the dimensions corresponding to the point positions and the velocities orientations, their distribution is uniform because the position and the direction of mobile object movement are arbitrary. Indeed, no information about the initial position or the orientation of mobile object movement is known. The velocity magnitude distribution is obtained directly from the empirical histogram of the observed data. Then, each time the region is centered on a different point X∈G, its distribution will change according to the dynamic points it contains. This search for the best region which will make it possible to identify the test group G as significant compared to the background model.

Thus, to detect and distinguish mobile objects among static objects, all the nodes in the binary tree as well as the space of the $\mathcal{H}$ regions are analyzed in order to evaluate the following hypotheses:

Hypothesis1: Any group of pixels which does not follow the random distribution of the background model is considered to be a group with independent movement. In order to obtain a quantitative value for the evaluation of this hypothesis, we use a measure called Number of False Alarms (NFA) as in [2] for each group in the binary tree. It is obtained by the following equation:

$$NFA(G) = N^2 * |\mathrm{H}| \min_{\substack{x \in G \\ h \in H \\ G \in H_x}} \mathrm{B}\left(N-1, n-1, p\left(H_x\right)\right) \qquad (1)$$

In this equation, N represents the number of elements of the initial vector of the data V, | H | is the cardinality of the regions and n is the number of elements in the test group G.

The term appears in the minimal function is the accumulated binomial law which represents the probability those at least n points including the point X (x, y, v_x, v_y) center of the region are inside the H_x region. A group G is said to be significant (it can correspond to a dynamic object on the scene) if NFA (G) $\leq$ 1. Then, a second evaluation taking into account only the significant groups will be carried out. Validation of the first hypothesis require a technique to distinguish the groups relate to static objects of the background model with the groups considered mobile by NFA. Hence, the utility of the M_{GK} technique.

B. M_{GK} concept

We consider the two patterns for a following association rule: let S be the static group in the binary tree; and either A the group considered mobile by NFA, where S, A ∈ h_X .

The intuitive meaning of an association rule $S \rightarrow A$ is as follows: "Whenever the pattern S appears, the pattern A also appears with a certain degree of assurance", or "any object that has the S pattern tends to also have the A pattern with an estimated degree of confidence". Therefore, to facilitate the interpretation of a rule (Figure 1), the normalization of the normalized measure associated with μ would be to reduce its values to the interval [-1.1] so that:

- -1 value corresponds to the incompatibility.
- Values strictly between -1 and 0 correspond to repulsion or negative dependence.
- 0 value corresponds to independence.
- Values strictly between 0 and 1 correspond to the attraction or oriented positive dependence.
- 1 value corresponds to the logical implication between the premise and the consequence of a rule S→A.

X and Y are two patterns for a data mining context. We define the measure M_{GK} by:

$$M_{GK}(S \rightarrow A) = \begin{cases} \dfrac{P(A'|S') - P(A')}{1 - P(A')}, & if\ S\ favors\ A \\ \dfrac{P(A'|S') - P(A')}{p(A')}, & if\ S\ disfavors\ A \end{cases} \quad (1)$$

For two not independent patterns S and A, two cases can occur: either there is mutual attraction, in which case the dependence is positive. Either there is repulsion, so there is a positive dependence between S and $\overline{A}$: this implies $S \rightarrow \overline{A}$ in the other hand, then between $\overline{S}$ and A: in this case we have $\overline{S} \rightarrow A$ on the other hand. In both cases, we will always have to consider a positive dependence. Then, decompose the measure MGK as follows:

$$M_{GK}(S \rightarrow A) = \begin{cases} M_{GK}{}^{f}, & if\ S\ favors\ A \\ M_{GK}{}^{d}, & if\ S\ disfavors\ A \end{cases} \quad (2)$$

In this case, the favorable component M_{GK}^{f} will guide the semantics of M_{GK}. These properties allow the M_{GK} quality metric to select fewer rules than the Confidence measure if we only use positive rules.

It makes it possible to jointly measure the difference in independence and the degree of statistical implication between two patterns.

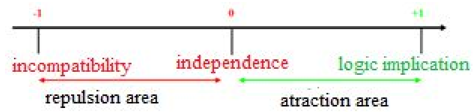


Figure 2. Distribution of probabilistic quality measure normalization values

Its coherence with the attraction and the repulsion between two patterns, it is less ambiguous and more intelligible than the χ_2 test of independence and the traditional Confidence. Moreover, the M_{GK} measure is favorably more discriminating than confidence.

C. M_{GK} and a contrario

Hypothesis 2: Any group that respects to the NFA criterion and validate by the M_{GK} measurement is considered to be a group that represents a salient object.

To answer this hypothesis, we calculate the distribution p composed of four independent distributions of each region h_X, which may contain a mobile or static group. After that, we have to calculate M_{GK} (S → A) and M_{GK} (A → S), then we compare the results obtained then we choose what is bigger and closer to 1.

To do this, we choose to use the formula of the favoring component of M_{GK}. The acceptance interval is between [0, 1].

$$M^{f}{}_{GK}(S \rightarrow A) = \frac{p_S(A) - p(A)}{1 - p(A)} \quad (4)$$

And

$$M^{f}{}_{GK}(A \rightarrow S) = \frac{p_A(S) - p(S)}{1 - p(S)} \quad (5)$$

Proposal: After the test, we take α as the final value of M^{f}_{GK}. Two cases are possible for validation:

- If α is between [0.95, 1], then we accept that groups that have a value NFA (G) $\leq$ α are accepted as mobile.
- Otherwise, we accept the first evaluation NFA(G)$\leq$ 1.

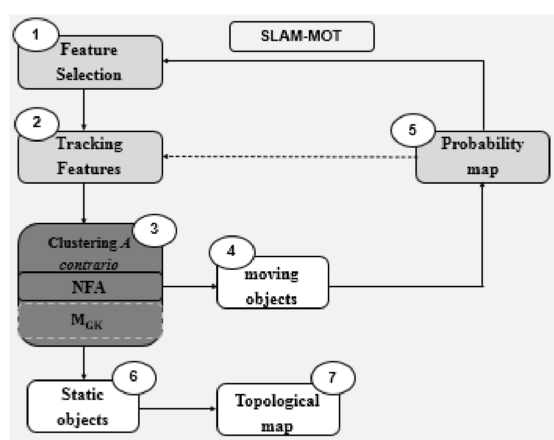


Figure 3. Diagram of our approach combined with SLAM.

At this threshold, the group G is considered mobile and is sent to the probability map which is used to track the dynamic objects in the next process. All groups that do not satisfy this condition are considered static. They will use for the construction of a topological map. Therefore, Figure 3 showed how to integrate our work in SLAM and MOT. The functions that should be added to implement the SLAM-MOT are the numbers 4 and 5.

D. Klt module

This module (light gray color in the Figure 3) is dedicated to the analysis of images acquired by the Smartphone camera. This one gives as result a set of points of interest characterized in R^4 obtained from the partial results of the functions:

- Feature Selection: Give the position (x, y) of the N best points of interest in the image.
- Tracking Features: Find the position (x, y) of the points in the next image and get their speed in the x and y directions (v_x, v_y).
- Probability map: Keep the cell position centered on each detected point of interest (x, y) in the image. A pixel value p_{ij} is assigned to each pixel in the cell according to a two-dimensional Gaussian distribution and its state over time. This map is reset every two tracking times.

These 3 functions are not performed for each image sequence. Feature Selection works only at the start of each tracking while the other two functions are executed for each image (from the second image for the Tracking features function).

Execution parameters: for each tracking, we use 150 points of interest to select in the image. The points found must be separated by at least 10 pixels. In order to select the points, which will be processed by the cluster module, these points must be tracking for at least 4 consecutive images and at the same time that the speeds v_x and v_y are greater than 1 pixel.

E. Cluster module

This module (gray color in the Figure 3) analyzes the points of interest characterized by the quadruplets (x, y, v_x, v_y), which give their respective positions and speeds along the tracking time. This one provides as result a set characterized by (x, y, v_x, v_y, C) where C represents the identifier of the group to which this point belongs. If C = 0, then the point is not part of an object with a coherent or defined movement. This module is executed at the end of each tracking module like as the Feature Selection function. On the other hand, the computation time of this one is a function of the number of points received at each execution. Using the "tick-tack" functions of the C identifier, the computation time is 1 ms to process 40 points but increases to a few seconds from 300 points received as input.

IV. EXPERIENCE AND RESULTS

We perform our experiment with a Smartphone, which has the following specifications: processor: Spreadtrum SC7731 – 1,3 GHz Quad Core, OS: Android 8.1, 1Gbyte of memory. The detection and tracking algorithm is tested on a sequence of 35 images taken by a Smartphone. We have the results below:

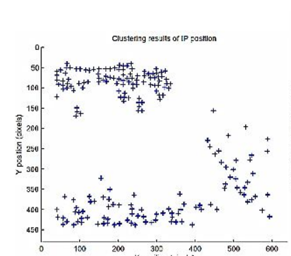
The points of interest accumulated in R^4 are represented in separate two-dimensional spaces. The x and y coordinates in the image are represented in pixels, the velocity magnitude in pixels/image and in degrees for the velocity orientation.

A. First case: Environment without mobile object

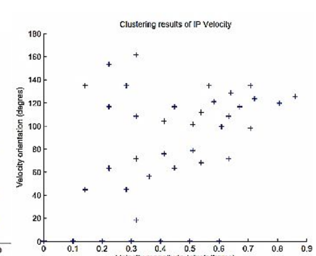
The first sequence of result shows two vehicles in the parking. This image is taking by a mobile user's camera. Pixel apparent movement is the result of the user's movement. In fact, we expect that the grouping method of dynamic points does not find coherent group. Figure 4a shows one of among 8 images used to accumulate the optical flow of points of interest, Figure 4b shows all accumulated point position and Figure 4c presents velocity magnitude and its orientation. In four dimensions (position and velocity) evaluations, the result shows that all data distribution is conformed to background model. In this case, tracking these objects is not necessary. In this first case, the user's movement speed is very low and no disturbance is present on the way of the user.

However, it is possible that the user is stumbled or some discontinuities on his way. This will have impacts on parasitic apparent movements during the acquisition of images, which requires a user motion compensation technique.

(a) Image initiale



(b) Grouping positions



(c) Grouping velocities

Figure 4. The grouping method evaluation with static targets

B. Second case: environment with rigid mobile objects

This experiment is focused on the detection of rigid moving objects. Figure 5 shows a scene where a car enters the field of view of the mobile user's camera. Initially, 150 points of interest are detected (shown in yellow in Figure 5a). Then, these points are tracking along 6 consecutive images. Figure 5b shows in blue the position on the image of all the accumulated points and in green the only group of mobile points identified as a dynamic object.

The position of these points corresponds exactly to the position of the ones on the car which enters the field of view. Figure 5c shows the magnitude and orientation of the velocities of the points.

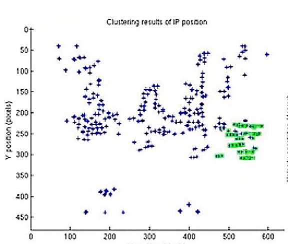
The green ones correspond to the detected object; all have the same orientation value since the orientation is around 0 and 360 degrees. Therefore, they correspond to the same direction.

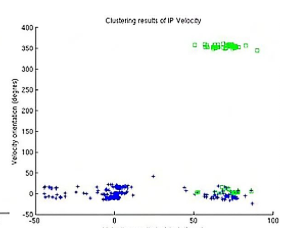
The time required for the detection of a dynamic object in the image is a function of the number of images used for tracking points, 6 in the case presented. The detection is done exactly even if there is a delay due to the detection of independent and coherent movements. Despite this, the detection of a rigid mobile object does not exceed 15 images after its first appearance.

C. Third case: environment with non-rigid mobile objects

Detection of dynamic objects becomes more complicated in the presence of non-rigid mobile objects (for example pedestrians) on the user's trajectory. For this test, we initially selected 150 points of interest, which are tracking for 20 consecutive images. The positions of the points as well as the two groups of dynamic points found are shown in Figure 6b.

(a) Initial image

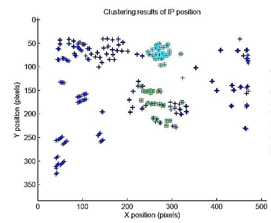


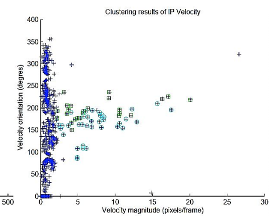


(e) Grouping positions (d) Grouping velocities

Figure 5. Representation of the positions and velocities of the points with a rigid moving object.

(a) initial image





(b) Grouping positions (c) Velocities grouping

Figure 6. Representation of the positions and velocities of the points in the presence of a non-rigid object.

Two groups are found even if there is only one pedestrian in the scene. The person's head and body are identified as a single object, shown in cyan, and the legs are detected as another object, which appears in green. Analysis of the result shows that the directions of movement of the points corresponding to the upper part of the body are different from those of the lower part. Moreover, the positions of these two groups in the image are not related due to the lack of points of interest on the trunk of the person, and the person proximity to the user's camera; this also prevents group merging.

Note that, in all the experimental results presented, the number of images used to accumulate points of interest is different. We note that to detect rigid objects, 8 images are sufficient. On the other hand, in the case of non-rigid objects, more images are necessary in order to properly represent the tracks. The pedestrian's case is more complicated due to movement of his feet.

V. CONCLUSION AND FUTURE WORK

The grouping technique presented in this article does not require any prior knowledge of the real scene, nor any prior information on the dynamic objects present in the scene. For this, we first used a scattered optical flow method by exploiting the KLT technique which allowed us to select and follow the moving points in order to distinguish it from static objects via the M_{GK} probabilistic measurement technique. Optimizing a grouping technique is useful for SLAM-MOT augmented reality applications. Compared to previous works, we were able to take a small step to solve the problem of speed of tracking objects on an uncontrolled

Copyright (c) IARIA, 2020. ISBN: 978-1-61208-779-5

scene because the sequences presented in this paper were acquired at 15 Hz, which is a time 4 s follow-up.

Despite these positive points, we found that our approach is sensitive for two situations: the first to the presence of a disturbance of movements during the user's movement. The second is occurred when detecting non-rigid objects. The latter may be due to the insufficient images accumulating to start the KLT algorithm. Consequently, if the points of interest followed are insufficient, the grouping algorithm cannot manage to group them in a precise manner.

In the future we plan to integrate a motion compensation technique to improve the disturbance due to camera movement. Measuring dissimilarity between correspondences could solve the problem of detecting and tracking deformable objects [14]. So, it will be useful to improve the false detection of points on non-rigid objects.

REFERENCES

[1] M. Ammar, S. L. Hégarat-Mascle, M. Vasiliu, and R. Reunaud, "An a-contrario approach for object detection in video sequence," international journal of pure and applied mathematics, Vol. LXXXIX No. 2, pp. 173-201, 2013.

[2] A. Gomez, A. Randall, G. Randall, and R. G. von Gioi, "A contrario 3d point alignment detection algorithm," IPOL journal image processing on line, vol. VII, pp. 399-417, 2017.

[3] N. Agarwal, C.-W. Chiang, and A. Sharma, "A study on computer vision techniques for self-driving cars," in international conference on frontier computing, pp. 629–634. Springer, 2018.

[4] A. Buyval, A. Gabdullin, R. Mustafin, and I. Shimchik, "Realtime vehicle and pedestrian tracking for didi udacity self-driving car challenge," in 2018 IEEE international conference on robotics and automation (ICRA), pp. 2064–2069, 2018.

[5] B. X. Chen, and J. K. Tsotsos, "Fast visual object tracking with rotated bounding boxes," in 2019 IEEE/cvf international conference on computer vision (ICCV) workshop, pp. 629-634, 2019.

[6] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in Proc. IEEE conference on computer vision and pattern recognition (CVPR 2000), vol. II, pp. 142–149, 2000.

[7] D. R. Feno, J. Diatta, and A. Totohasina, "A basis for the association rules of a valid binary context within the meaning of the mgk quality measure," in Proc. of the 13`eme rencontre de la société francophone de classification, pp. 105-109, Metz, France, 2006.

[8] R. Giraud, and Y. Berthoumieu, "Texture Superpixel Clustering from pathch-based nearest neighbor matching," 2019 27th european signal processing conference (EUSIPCO), pp. 1-5, june 2019.

[9] Q. Guo, W. Feng, C. Zhou, C.-M. Pun, and B. Wu, "Structure-regularized compressive tracking with online data-driven sampling," IEEE transactions on image processing, 26 (12), pp. 5692–5705, 2017.

[10] Y. Hua, K. Alahari, and C. Schmid, "Online object tracking with proposal selection," in proceedings of the IEEE international conference on computer vision, pp. 3092–3100, 2015

[11] M. Kristan, A. Leonardis, J. Matas, M. Felsberg, R. Pflugfelder, L. Čehovin Zajc, T. Vojir, G. Häger, A. Lukežič, A. Eldesokey, and G. Fernandez et al,"The seventh visual object tracking VOT2019 challenge results," in international conference on computer vision (ICCV) workshop, pp. 639-654, 2019.

[12] Y.-G. Lee, Z. Tang, and J.-N. Hwang, "Online-learning-based human tracking across nonoverlapping cameras," IEEE transactions on circuits and systems for video technology, 28(10), pp. 2870–2883, 2017.

[13] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," In Proc. DARPA Image Understanding Workshop, Pp. 121–130, April 1981.

[14] G. Nebehay and R. Pflugfelder, "Clustering of static-adaptive correspondences for deformable object tracking," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2784–2791, 2015.

[15] L. Rout, D. Mishra, R. K. S. S. Gorthi, and Siddhartha, "Rotation adaptive visual object tracking with motion consistency," in 2018 IEEE winter conference on applications of computer vision (WACV), pp. 1047–1055, 2018.

[16] J. Shi and C. Tomasi, "Good features to track," in Proc. IEEE international conference on computer vision and pattern recognition (CVPR 1994), pp. 593–600, june 1994.

[17] Q. Wang, L. Zhang, L. Bertinetto, W. Hu, and P. H. Torr, "Fast online object tracking and segmentation: a unifying approach," 2019 IEEE conference on computer vision and pattern recongnition (CVPR), pp. 1328-1338, 2019.

[18] R. Xu, S. Y. Nikouei, Y. Chen, A. Polunchenko, S. Song, C. Deng, and T. R. Faughnan, "Realtime human objects tracking for smart surveillance at the edge," in 2018 IEEE international conference on communications (ICC), pp. 1–6, 2018.

[19] D. Prasad, C.K. Prasath, D. Rajan, and L. Rachmawati, "Object Detection in a Maritime Environment: Performance Evaluation of Background Subtraction Methods," in 2018 IEEE Transactions on Intelligent Transportation Systems, pp. 1-16, 2018.

[20] S. S. Mohamed, N. Tahir, and R. A. Adnan, "Background modelling and background subtraction performance for object detection," in 2010 Signal Processing And Its Applications (CSPA), pp. 236-241, 2010.