Search and Automatic Summarization of Relevant Features in Forensic Scenarios

Erick V. C. L. Borges, Igor L. P. Andrezza,
Rajiv A. T. Mota
Vsoft Tecnologia
João Pessoa, PB, Brazil
{igor.andrezza, erick.borges,
rajiv.mota}@vsoft.com.br

Leonardo V. Batista
Centro de Informática - CI
Universidade Federal da Paraíba - UFPB
João Pessoa, PB, Brazil
leonardo@ci.ufpb.br

***Abstract*—Methods and tools that perform the analysis and the automatic monitoring of environments are increasingly needed due the growing presence of video recording devices. This work aims to extract some relevant features of forensic scenes, performing summarization only in the relevant data. The proposed methods aim to detect and analyze motion, face detection and recognition, perform the tracking of human faces and recognition of predominant color in the clothing of individuals. The results are comparable to the ones found in the literature and may contribute to the fast extraction of information needed for human analysis, to assist in the interpretation and documenting the results.**

***Keywords*-Forensic Computing; Video Summarization; Computer Vision.**

I. Introduction

Increasingly common in public and private sectors, video recording systems are often used in criminal investigations and analysis. The ability to extract information from videos has potential to assist investigations, help solving possible doubts arising when it is necessary analyze and to solve cases in civil and criminal matters. Despite the great amount of information, the acquisition is a big challenge when manually made by human beings, as the capacity of the human visual system of inspecting an accelerated video is usually limited and depending on the level of activity in scene, it may be an extremely slow and heavy process [1].

Thus, the understanding and automatic interpretation of activities performed by humans is quite valuable to aid investigators, directing their attention to the potentially useful content [2]. The automatic interpretation may aid the investigations in several ways, as better trace analysis, survey of important data, case documentation, continuous monitoring to prevent crimes and accidents, real-time alerts to authorities about possible events, etc.

To automatically interpret the activities performed by humans, tasks such as detection, recognizer, tracking should be handled accurately and efficiently. An issue related to this subject is that sometimes the evidences are hidden in a chaotic environment, increasing the level of difficulty to interpret the scene, compared to a constrained environment. Another factor that can influence the problem resolution is the quantity and quality of available data in the scene. Low quality images, mostly due to practical and economic issues, aggravate these situations.

The purpose of this work is to develop an support tool to relevant information extraction of forensic scenes, anticipating and automating tasks which would be made by the examiner and summarizing the data of interest in a forensic scene. The methods covered by this tool aim to extract information related to:

1) *Movement detection*, where objects of interest are detected and information about them are extracted;
2) *Human face*, detection and identification used to find possible suspects;
3) *Predominant color on individual's clothing*, used to identify features found in suspects.

The examiner can utilize any proposed method to search or summarize features of your interest.

II. Related Work

Low-quality videos recorded from surveillance cameras can hide some important information, because of that, some authors aim to improve those images instead of extracting some information from them. The work of Jerian *et al.*[3] is an example of that; in this work, they present a software environment, named Modular Image Processing Environment (MIPE) which has the goal of providing to the examiner the knowledge and full control over the enhancement methods on image that can improve the images quality.

Surveillance systems are inclined to introduce an active and safe intelligence through computational methods able to extract information from scenes automatically. A practical and challenging approach is to detect and locate abnormalities, low occurring events, so the system warns that unexpected activities are on going [4] [5]. Some techniques are employed to extract many different information from surveillance videos, as intrusive objects movement [6], abandoned objects classification [7], traffic analysis [8], people tracking [9], and others.

Some tools which use feature extraction methods and have similar goals were proposed in literature, authors highlight the necessity of intelligent monitoring systems that aid processing and interpreting information and proposed a tool aiming allow searching through large amounts of data [1] [10] [11]. Those tools are able to detect and track as human as vehicle activities, find people matching a profile, rebuild high resolution images, detect suspects through background removal, track suspects, extract appearance features like height, speed and body parts, identify accessories and skin color, analyze suspects activity and make safe decisions.

III. MATERIALS

Some image databases were used to perform tests and evaluate the system.

Yale Face Database A [12], Extended Yale Face Database B [12] and Vsoft-faces [13] are the datasets chosen to develop the face recognition method, because it is utilized by others found in literature. In order to perform tests with videos resembling to surveillance cameras, an own video database was made, with the same individuals present on Vsoft-faces database [13].

The Incremental Learning for Visual Tracking Dataset [14] was used to perform tests on the face tracking method, an indoors and an outdoors video were extracted.

The Change Detection Benchmark Dataset 2012 [15] provided videos from the baseline category to evaluate the movement detection and analysis method. This category was chosen because the videos have suitable features for this particular method.

IV. PROPOSED METHODS

The algorithms on this paper cover three main subjects: movement, human face and clothing.

First is the motion detection and analysis, comparing consecutive frames and detecting substantial differences between them, and after classifying the motion according to established criteria. The second element extracts relevant information from a face image. Begins with face detection on a frame followed by identification of the individual through face analysis and verifying whether he is present in the database or not. The third component estimates the position of the individual's body based on the detected face to extract information of predominant color in his clothing, using the HSV color system to make sure the system get the correct information about color. At the end, a summarization was proposed, where the examiner chooses what to extract from a scene, dispensing unnecessary data.

A. Movement Detection and Analysis

To extract information about motion in a scene, the main idea was the background subtraction. It avoids false detection through background update, as if it was took static, every permanent change in the scenario would be treated as movement. This approach was chosen because surveillance cameras are usually statically placed, so the background is always well established.

The proposed method updates the background whenever the number of pixels detected as movement is less than a predefined threshold, chosen by the user. This background update is achieved by using the following equation:

$$bg = \begin{cases} cf, & \Delta I <= \tau \\ bg, & \text{otherwise} \end{cases} \tag{1}$$

where ΔI is the quantity of pixels classified as movement on the frame, *bg* is the current background, *cf* is the current frame and τ is the threshold established by the user.

Aiming to make the detection even more specific, some parameters refine the search by size and predominant color. To find an object of specific size the pixels from the background are subtracted from the actual frame and undergo to blobs detection, which is grouping these pixels in objects and distinguish the group of pixels in the frame. After separating the objects, width, height and area are measured, allowing to be filtered by the examiner. To identify the predominant color, the image is converted to the HSV color system and then the algorithm identify the color represented in the largest number of pixels belonging to that object.

The data can be plotted in a chart showing motion over time, so periods with more movement than normal can be pointed out. Figs. 1 and 2 show some features of the system.

Fig. 1. Frame pointed as having the most intensity of motion (plane crash).

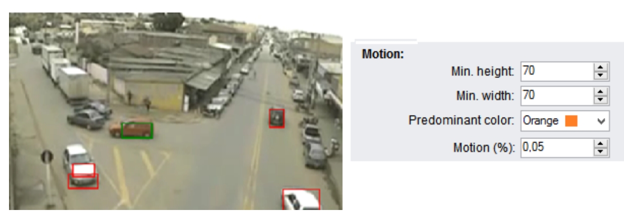

Fig. 2. Search for orange moving objects, 70x70 pixel size minimum.

This is a very important feature because allows the examiner to focus on the frames of the video which really present some movement or present an amount of movement bigger than the chosen threshold.

B. Human faces detection, recognition and tracking

The face detection step used the Viola-Jones [16] algorithm, which is based on pattern recognition and is already provided by the multiplatform library OpenCV. Previous tests defined the parameters set to locate the faces, like scale factor and face size.

After the detection, some normalization techniques are applied aiming to improve the extracted face images. First, geometric normalization filters pose variation, this algorithm uses face landmarks detected by ASM [17]. Using the eye center locations, the image is rotated to remove the slope between the eyes. Then, the image is scaled to a standard predetermined size.

Sequentially, illumination normalization filters shades, contrast and light variations. The methods used were Histogram Equalization, Non-Linear Transform (Log) and LogAbout [18]. Fig. 3 shows the flow of this step and the result of each method.

(a) Original (b) Rotated (c) Hist. Eq. (d) Log (e) LogAbout

Fig. 3. Face normalization step.

The face recognition is performed using Local Binary Pattern Histograms (LBPH) [19] to extract texture features from the face image. The normalized image is the input for the feature extraction step. The image is then divided into blocks which are analyzed to calculate the Local Binary Pattern. Afterward, the set of LBP is concatenated to compose the features vector, which is a collection of LBP histograms. (see Fig. 4)

Aiming to classify features vector, the classification stage computes a score for each pair of vectors through a matching metric. In this proposed method, the used matching metric is defined by the following equation:

$$d(H_1, H_2) = \frac{\sum_{i=0}^{I}(H_1(i) - \overline{H}_1)(H_2(i) - \overline{H}_2)}{\sqrt{\sum_{i=0}^{I}((H_1(i) - \overline{H}_1)^2 \sum_{i=0}^{I}((H_2(i) - \overline{H}_2)^2}} \quad (2)$$

$$\overline{H}_k = \frac{1}{N}\sum_{0}^{i} H_k(j) \quad (3)$$

where, H_1 is the first input of the matching process, H_2 is the second one and N is the number of values in the histogram. Fig. 4 illustrates the proposed methods to face recognition.

At the face tracking step, the algorithm compares each found face with faces in previous frames. First, the position is compared, if the face is placed at a shorter distance than a certain threshold related to the previous one, they are taken as the same individual, otherwise it is taken as a new face. Since the face is treated as new by this step, the algorithm verify if the face belongs to the database. If the verification is positive its ID is assigned, if not, the system adds the face to the database and assign a new ID.

The query to the database is performed by the previously referred method, so, every new face generates a LBP histogram and then the algorithm look for a similar one.

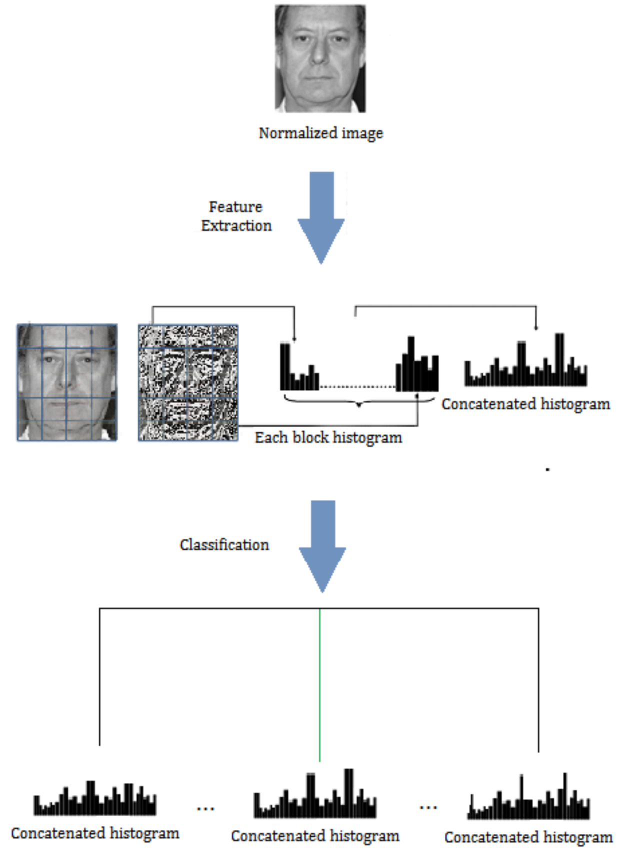

Fig. 4. Face classification step.

C. Clothing's Predominant Color Recognition

The information about clothing's color is useful in many applications, finding this color is important, e.g., when the examiner wants to look for a suspect according to witnesses.

The proposed method uses the face localization to presume the clothes location and so, to find out the main color in the person's clothing. Finding the region where supposily, the clothing information is placed, the method presumes the person is standing and, because of that, the information is located under the face. The method looks for a region under the face wich has a relative size to the face and can represent the subject's clothing. After finding this region, the method converts the pixels into it to the HSV color system. The conversion is necessary because the HSV color system is invariant to white lighting high intensity and ambient lighting, so it can be a good choice for color detection methods. The resulting image is analyzed to identify predominant color in the clothing region, previously found.

D. Scenes Summarization Through Developed Methods

A set of search criteria can be used as filters to extract relevant information. The summarization proposed by this paper allows real-time analysis or process the whole video storing relevant data for compiled analysis.

A proposal of summarization is presented after the development of the methods presented previously. The summarization has a set of parameters for each method used, so it is possible to analyze the scene and extract only desired information. The summarization used in this work allows the user to perform a real-time analysis of the scene or store the data for a posterior analysis.

The application allows the user to choose the set of criteria for the searching, like the size of the objects, predominant color of the objects, the intensity of the movement of the object, face size, clothing main color and the identity of the searched subject. After finding the frames wich satisfy the defined criteria, the developed application allows the storage of these frames on a new video, reducing the amount of data to be analyzed and makes easier the posterior analysis. The application also allows to storage the frame wich has the highest level of movement and can generate a plot associating the level of movement in the scene to the timing.

The application can generate a report which contains the summary of all extracted information from the scene too. This report can be useful to storage the data, visualize the information easier and substantiate the analysis process. The report assists in the interpretation of the analysis and a possible legal argumentation.

The choice of the criteria is made through a graphic interface shown in the Fig. 5 which is a print screen of the developed application. The Fig. 6 shows a screenshot of the developed application where a scene is analyzed.

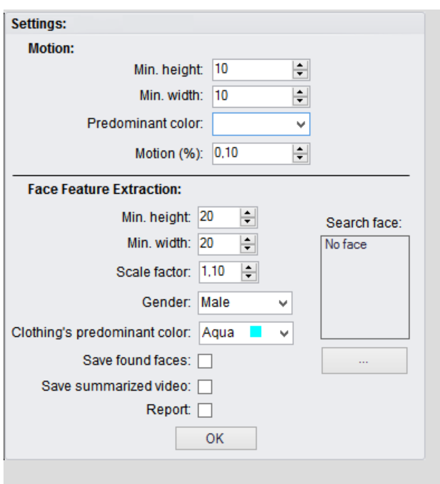

Fig. 5. Screenshot of the developed application settings.

V. Experimental Results

Tests were made according to each method's objective.

A. Movement Detection and Analysis

In order to establish a comparative with methods found in the literature, some metrics specified in the Change Detection Benchmark Dataset 2012 [15] were measured. This was done comparing every pixel of the movement frame with the ground truth presented by the database. Fig. 7 shows this process.

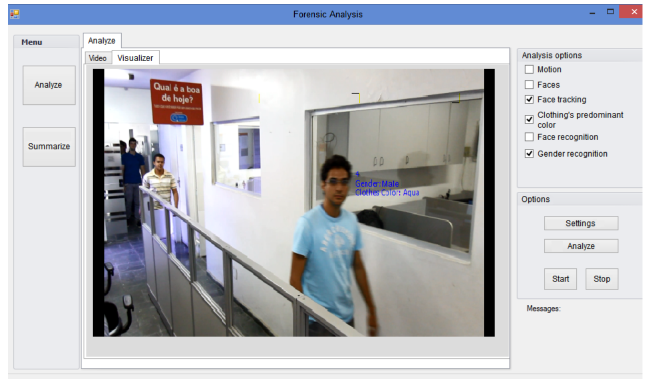

Fig. 6. Screenshot of the developed application.

(a) Original frame (b) Method (c) Ground truth

Fig. 7. Comparative between the movement frame provided by the proposed method and the ground truth of the *Change Detection Benchmark Dataset 2012* [15].

All videos of the baseline category were used. For every video were measured the rates of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). After, were measured the following metrics: Specificity (Sp), Sensibility or Recall (Re), False Positive Rate (FPR), False Negative Rate (FNR), Percentage of Wrong Classifications (PWC), Precision (Prec) and F-Measure.

$$Re = \frac{TP}{(TP + FN)} \tag{4}$$

$$Sp = \frac{TN}{(TN + FP)} \tag{5}$$

$$FPR = \frac{FP}{(FP + TN)} \tag{6}$$

$$FNR = \frac{FN}{(TP + FN)} \tag{7}$$

$$PWC = \frac{100 \cdot (FN + FP)}{(TP + FN + TN + FP)} \tag{8}$$

$$Prec = \frac{TP}{(TP + FP)} \tag{9}$$

$$F - Measure = \frac{2 \cdot Prec \cdot Re}{(Prec + Re)} \tag{10}$$

The results provided by the proposed method are consistent with other modern methods [20] [21] which were evaluated under the same circumstances. Results are disposed in Table I.

TABLE I
METRICS OF THE BASELINE CATEGORY FROM THE CHANGE DETECTION BENCHMARK DATASET.

Metrics	**Proposed Method**	**Yoshinaga *et al.***	**Strauss *et al.***
Re	0.6303	0.7551	0.7044
Sp	0.9932	0.9940	0.9923
FPR	0.0068	0.0060	0.0077
FNR	0.3697	0.2449	0.2956
PWC	2.3570	1.9154	2.2142
Prec	0.8001	0.7833	0.7009
F-Measure	0.7029	0.7554	0.6616

Analyzing the results, it is clear that FNR and Re rates are an issue of the proposed method. False negatives are quite influential, these are found when the algorithm does not assign motion to an area where it happened. Besides, some images contain noise counted as false positives, which may have influenced some metrics.

B. Detection, Recognition and Tracking Human Faces

Feature vectors were extracted and compared on an approach 1:N (one to many), under the cross validation concept, which divides the database in groups and, for each stage, one group is used to training and the others are used for testing. The cross validation ends when every group is chosen as training once. Each match between vectors provides a score through the similarity metric, then the input vector is assigned to its most similar training vector. Table II shows the results from this experiments.

TABLE II
SUCCESS RATE OF FACE RECOGNITION.

Geometric Normalization	**Illumination Normalization**	**Vsoft**	**YALE A**	**YALE B**
No	None	100%	96,7%	99,6%
	Log	94,9%	81,1%	100%
	LogAbout	93,6%	76,6%	100%
	Histogram	100%	76,6%	100%
Yes	None	70,5%	81,1%	83,3%
	Log	70,5%	82,5%	84,6%
	LogAbout	75,6%	77,9%	85,4%
	Histogram	69,2%	79,9%	83,7%

Examining the results, it is evident that the proposed method had higher success rates without geometric normalization. Analyzing the process, it was found that some images were normalized based on wrong information about the eye location provided by the ASM, which proved to be not accurate enough to this task. Aiming to evaluate more precisely the proposed recognition method, this kind of error had to be removed, so the images from YALE A dataset were all manually marked. After, the normalization step was remade with the right eye locations in all images. Table III shows the results of the new normalization and compare to the old values.

Performing the geometric normalization with handmade eye location made the algorithm reach higher success rates than using ASM points. It shows that if the user has an efficient way to find the coordinates of the eyes, the proposed method can improve the success rates.

It was also accomplished experiments aiming to test the accuracy of the proposed methods on images extracted from videos. The images from the VSOFT-Faces dataset [13] were used for training and the faces extracted from videos which have the same subjects of the dataset were used for testing. The Table IV shows the results of the proposed methods about videos faces.

The Table IV presents a success rate lower than the tests performed on images from other datasets, this might be explained by the fact of face images extracted from the video records have a higher variation of pose and background, and also have a blur effect due the movement of the subjects.

Regarding the face tracking on the video, for each new face detected a new search in the system is performed in order to discover if the face was already registered. If the search fails, the face is registered as a new subject in the system, otherwise, the face is labeled as being of the same subject of the register found in the process.The identification assigned to each face is compared to a ground-truth, through those comparations are calculate the hits and mistakes to evaluate the system.

The Table V shows the success rate of the identification tests performed on the video records..

One of the problems faced by the proposed method in the performed tests was to cope with the variations of conditions of the same subject in different moments on the video record, like the usage or not of glasses, pose and lighting variations. Because of those variations, in some cases, the proposed method was not able to find the subject among the registers in the system. So, this behavior increased the number of registered subjects in the system.

C. Clothing's Predominant Color Recognition

Own videos were used to perform the test, recorded with individuals of Vsoft-Faces database [13]. Likewise, videos found in literature as Incremental Learning for Visual Tracking [14] were also used, choosing images containing at least one face. The face tracking was applied in every video and a small ground truth about the predominant color of everyone's clothing was manually made to evaluate the success rate of the method. Then, each attribution by the system was compared to the ground truth, if they match it is a success, otherwise it is an error.

In every test the success rate was measured and disposed in Table VI. The average success rate in this case is 81%.

Analyzing the method, some limitations can be pointed out, e.g. variations of lighting, where the values of saturation and

TABLE III
SUCCESS RATE OF FACE RECOGNITION WITH MANUALLY LOCATION OF THE EYES FOR GEOMETRIC NORMALIZATION, ON YALE A DATASET.

Illumination Normalization	**ASM points**	**Handmade Location**
None	81,1%	96,1%
Log	82,5%	98,0%
LogAbout	77,9%	92,2%
Histogram	79,9%	96,7%

TABLE IV
SUCCESS RATE OF FACE RECOGNITION IN FACES CAPTURED BY VIDEOS.

Illumination Normalization	Sucess Rate
None	61,82%
Log	59,32%

TABLE V
SUCESS RATE TO FACE REIDENTIFICATION

Video	Faces Count	Handmade Individuals Differents Found	Automatic Individuals Differents Found	Sucess Rate
Own video record 01	164	14	13	90.96%
Own video record 02	1572	13	21	61%
Outdoor sequence [14]	315	1	1	88.88%
Indoor sequence [14]	407	1	3	59.00%

bright prevented the system to find the color purity, and shirts with color variations, like stripes or patterns.

VI. CONCLUSIONS

The automatic interpretation of forensic scenes is becoming increasingly necessary because the large amount of data makes the analysis extremely slow and often humanly impossible. Therefore, scene analysis technologies to inspect the content and automatically extract features are getting even more relevant to daily life. Computational methods which analyze and interpret movement in scenes and are able to extract information from the presence of human beings or even recognize human features are particularly useful in many scenarios.

The overall goal of this work was development and evaluation of methods for analysis of relevant data in scenes, videos or images, extracting information to aid forensic procedures. The methods were successfully developed into an application that aid examiners with forensic scenarios.

Regarding to movement detection, the method has proven itself efficient and comparable to other ones found in literature. An artifice that may improve the evaluated metrics is a postprocessing to improve the object segmentation.

Despite of reaching reasonable success rates, some weaknesses can be pointed out on clothing's predominant color recognition. As a future work, it is proposed to improve the segmentation of the body and to develop a strategy to deal with striped shirts or clothes with prints. These improvements will allow the system to make more precise descriptions about the subject's clothing.

Regarding to face recognition, with an accurate pupil detector, the LBP based classifier together with normalization techniques can be used for frontal face recognition, as it reaches success rates above 97%. It is proposed, as a future work, to try other versions of the LBP and other matching methods to compare the LBPH outputs, instead of the correlation method. It is also proposed to look for methods which can detect facial landmarks more precisely and correct the geometric normalization.

There are still big challenges in video summarization in unconstrained environments, largely due to lighting variation as well as occlusion and mixed background. Nevertheless, the method and summarization presented by this paper can assist analyzing videos focusing the examiner in points of interest, which may be helpful to speed the process. The involvement of the examiner is crucial because it makes the investigation a hybrid analysis, combining the expertise of the examiner with the capabilities of the computational system.

TABLE VI
CLOTHING COLOR RECOGNITION.

Video	Frames	Individuals found	Success rate
Own video record 01	1604	13	93%
Own video record 02	4542	21	84%
Outdoor sequence [14]	569	6	50%
Indoor sequence [14]	770	2	100%

REFERENCES

[1] J. Thornton, J. Baran-Gale, D. Butler, M. Chan, and H. Zwahlen, "Person attribute search for large-area video surveillance," in *Technologies for Homeland Security (HST), 2011 IEEE International Conference on*. IEEE, 2011, pp. 55–61.

[2] W. R. Schwartz, "Aplicação de observação de pessoas em computação forense."

[3] M. Jerian, S. Paolino, F. Cervelli, S. Carrato, A. Mattei, and L. Garofano, "A forensic image processing environment for investigation of surveillance video," *Forensic science international*, vol. 167, no. 2, pp. 207–212, 2007.

[4] T. Xiao, C. Zhang, and H. Zha, "Learning to detect anomalies in surveillance video," *Signal Processing Letters, IEEE*, vol. 22, no. 9, pp. 1477–1481, 2015.

[5] A. A. Sodemann, M. P. Ross, and B. J. Borghetti, "A review of anomaly detection in automated surveillance," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 42, no. 6, pp. 1257–1272, 2012.

[6] C. H. Chen, T. Y. Chen, Y. C. Lin, and W. C. Hu, "Moving-object intrusion detection based on retinex-enhanced method," in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2014 Tenth International Conference on*. IEEE, 2014, pp. 281–284.

[7] A. F. Otoom, H. Gunes, and M. Piccardi, "Feature extraction techniques for abandoned object classification in video surveillance," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE, 2008, pp. 1368–1371.

[8] V. Katkar, S. Kulkarni, and D. Bhatia, "Traffic video classification using edge detection techniques," in *Nascent Technologies in the Engineering Field (ICNTE), 2015 International Conference on*. IEEE, 2015, pp. 1–6.

[9] J. Li, R. Li, R. Mohedano del Pozo, B. Adán, C. Cuevas Rodríguez, G. Gallego Bonet, N. García Santos *et al.*, "Real-time surveillance application by multiple detectors and compressive trackers," 2015.

[10] L. Attard and R. A. Farrugia, "Vision based surveillance system," in *EUROCON-International Conference on Computer as a Tool (EUROCON), 2011 IEEE*. IEEE, 2011, pp. 1–4.

[11] S. Aramvith, S. Pumrin, T. Chalidabhongse, and S. Siddhichai, "Video processing and analysis for surveillance applications," in *Intelligent Signal Processing and Communication Systems, 2009. ISPACS 2009. International Symposium on*. IEEE, 2009, pp. 607–610.

[12] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 6, pp. 643–660, 2001.

[13] A. Marinho, "Uma nova versão de um sistema de detecção e reconhecimento de face utilizando a transformada cosseno discreta," 2012.

[14] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 125–141, 2008.

[15] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "Changedetection. net: A new change detection benchmark dataset," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*. IEEE, 2012, pp. 1–8.

[16] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[17] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[18] H. Liu, W. Gao, J. Miao, and J. Li, "A novel method to compensate variety of illumination in face detection." in *JCIS*. Citeseer, 2002, pp. 692–695.

[19] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *Computer vision-eccv 2004*. Springer, 2004, pp. 469–481.

[20] S. Yoshinaga, A. Shimada, H. Nagahara, and R. Taniguchi, "Background model based on intensity change similarity among pixels," in *Frontiers of Computer Vision,(FCV), 2013 19th Korea-Japan Joint Workshop on*. IEEE, 2013, pp. 276–280.

[21] O. Strauss, D. Sidib, and W. Puech, "Quasi-continuous histogram based motion detection," Technical Report, LE2I, Tech. Rep., 2012.