A Shadow Removal Approach for a Background Subtraction Algorithm

Rosana Barbuzza[1,3], Leonardo Dominguez[1,2], Alejandro Perez[1,3], Leonardo Fernandez Esteberena[1,2], Aldo Rubiales[1,3], and Juan P. D'amato[1,2](✉)

[1] PLADEMA, Universidad Nacional del Centro de la Provincia de Buenos Aires, Buenos Aires, Argentina
juan.damato@gmail.com
[2] National Council Scientific Technical Research, CONICET, Buenos Aires, Argentina
[3] Comisión de Investigaciones Científicas, CICPBA, Buenos Aires, Argentina

Abstract. This paper presents preliminary results of an algorithm for shadow detection and removal in video sequences. The proposal is that from the base of the background subtraction with the Visual Background Extraction (ViBE), which identifies areas of movement, to apply a post processing to separate pixels from the real object and those of the shadow. As the areas of shadows have similar characteristics to those of the objects in movement, the separation becomes a difficult task. Consequently, the algorithms used for this classification may produce several false positives. To solve this problem, we set to use information of the object involved such as the size and movement direction, to estimate the most likely position of the shadow. Furthermore, the analysis of similarity between the present frame and the background model are realized, by means of the traditional indicator of normalized cross correlation to detect shadows. The algorithm may be used to detect both people and vehicles in applications for safety of cities, traffic monitoring, sports analysis, among others. The results obtained in the detection of objects show that it is highly likely to separate the shadow in a high percentage of effectiveness and low computational cost; allowing improving steps of further processing, such as object recognition and tracking.

Keywords: Video processing · Object detection · Segmentation

1 Introduction

Video analysis systems have become a useful tool not only in industry but also in the research field. Digital video surveillance has proved fundamental for both forensics pericia and crime prevention. Particularly, security in Argentina is a critical issue which must be dealt with immediately; therefore, many monitoring centers and camera observers have arisen. Being this procedure not efficient

A. E. De Giusti (Ed.): CACIC 2017, CCIS 790, pp. 101–110, 2018.
https://doi.org/10.1007/978-3-319-75214-3_10

enough, it is highly important to count with techniques of video analysis that may help operators with their daily tasks. Consequently, this research group has submitted different research works mainly showing the architecture of an open distributed system, and also scalable [1].

The complexity of the presented algorithms lies in working in dynamic environments, as exterior cameras are, being under the threat of weather conditions. To diminish these hardships, moving windows as detailed in [2], are usually used, being their objective focusing on applying a logic operation to process only those movements detected within their range of vision.

These object tracking algorithms are eventually used in video-analysis platforms to detect and analyze certain situations of special interest. In this context, shadows considerably affect the perception of the detected objects, as shown in Fig. 1, since it drastically alters an aftermath classification by color or size; then, it has to be reduced someway.

Fig. 1. Images of different cases were the shadow changes the object size and form.

The present research work focuses on being able to detect and eliminate shadows from videos in grayscale, or color space transformed to grayscale (i.e. rgb2gray o HSV corresponding to component Value). This consideration is proper since many low-cost cameras do not usually provide images of sufficient quality, also affected by video stream compression. The novelty of the suggested method is that it uses information coming from the object and its orientation, to determine before hand the position of the shadow. Then, a traditional method is applied in the external region of the object which considers characteristics of texture and color in areas of shadows.

This research work is organized as follows: Sect. 2 describes states of the Arts; Sect. 3 the VIBE method for background subtraction and the introduced modifications. In Sects. 4 and 5, the proposal shadow detection method and the results are presented. Finally, Sect. 6 presents conclusions and future works.

2 State of the Art

The techniques of detection of events used in Video-Survillance systems are mainly based on quickly discriminating of movement from a fixed video camera. Research works as [3], describe the most common algorithms to detect and track objects. In [4], there is a special comparison between the different algorithms of basic detection, concluding that their combination may be absolutely useful to diminish false positive rates; while keeping the time and rate of true positives.

The background subtraction commonly used classify shadows as a part of the objects, what alters both size and shape of the objects, affecting consequently the efficacy of such algorithms [5]. Also, it hinders some other procedures which need the result of the detection of objects, such as classification or tracking of the trajectory and also, analysis of certain behavior: loitering, vandalism, traffic lawbreaking, among others. This problem also affects those techniques based on characteristics [6].

There is different research on the detection and separation of shadows. Some algorithms that worked with static images are computationally complex and are not applicable for video analysis in real time [7]. Other algorithms are specially designed for video in either greyscale or color, and also have a lower computational cost. To perform classification, most bases on the basic common characteristics of the area with shadows: areas darker than the background Of the scene, uniform and invariable color or texture, etc. Even though these characteristics allow creating candidates to determine the areas of shadows, they are not decisive; therefore, all of the methods fail in the classification because many times part of the objects also satisfy these conditions [8,9]. In specific, another group of algorithms include information about the geometry of the shadow or the illumination model, some of them only specialized in the shadows of people.

3 Background Subtraction in Video

The present research work makes use of the Visual Background extractor (ViBE), as suggested in [10], which behaves properly in different environments commonly used in video surveillance. Among the virtues in applying this method we may highlight the low computing time, the high detection rates and the strength in noise existence, highly necessary features in surveillance camera captures used at present. Likewise, the present proposal may be also applied to some other algorithms [3]. Figure 2 depicts background subtraction accomplished with ViBE for a particular frame of urban video surveillance. In the example, the original image

Fig. 2. Video image, left, and background subtraction with ViBE, right, for a video captured in a video surveillance camera.

is shown and to its left, the foreground components detected, people and vehicles. The shadow projection coming from the bus is part of the object, notably deforming its shape, and increasing its size to approximately a 50%.

ViBe is a pixel based modeling background algorithm, choosing at random way samples of the intensity of each pixel considering the previous frames from the incoming video. Figure 3 depicts the sequence of steps of the original ViBe algorithm and the modules included for a better classification. The module Detection, classifies scenes as background or foreground, calculating the distance between each pixel in a present image with regard to the samples stored in the background model.

Updating of the background model for each pixel is aleatory, being likely to replace one of the stored values for a new intensity value of the same pixel in the present frame (Fig. 3, module Update model). This updating mechanism allows preserving background pixels, while motion ones are partially discarded in time. It is essential that only those pixels processed and classified as background replace samples in the corresponding background model. Inserting wrongly classified pixels which belong to objects in motion, foreground, or that become uncertain in the classification may significantly alter the detection results. This instance is of utmost importance since it has allowed the introduction of improvements to the original ViBE method achieving a better rate detection, mainly to adapt the algorithm to scene movements or dynamic background, detected as false positives, or the refilling of those objects not fully detected, false negatives [11,12].

This paper adds the algorithm of morphologic operation such as opening, closing, holes refilling. In order to attain connected components, the larger objects are selected whereas the little and isolated ones are discarded (Countour and Refilling module). Once this process was carried out, the algorithm of cast shadow separation of the real object is applied (Shadow module).

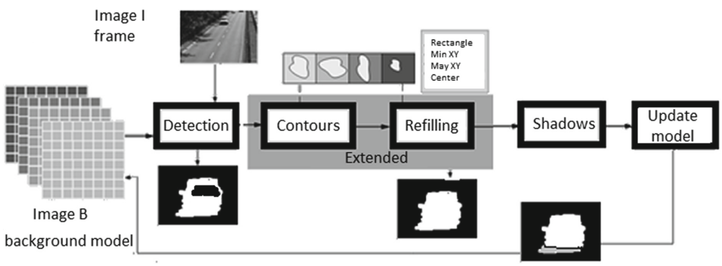

Fig. 3. ViBE algorithm and implementation of modules of improvement

The present module suggests correcting only those pixels classified as foreground which correspond to shadows. It should be noticed that it is not proper to classify shadows as background, since these samples should not modify or alter the background model, when updating the model, to avoid further errors when classifying.

4 Shadow Detection Proposal

To detect and eliminate shadows it is possible to apply the Normalized Cross Correlation NCC between the present image, I, and an image representative of the background, B, adjusting with properties typical of a shadow as shown in [13] and also, incorporating to the original NCC method, elemental knowledge of the objects to improve detection rate, both detailed below.

4.1 Identification of Potential Pixels with NCC

NCC indicator allows identifying similar images with different scales of intensity. To represent the background, B, an image obtained from those samples randomly selected by ViBE as representative of the background model was used, see Fig. 3.

For each pixel (i, j) classified as foreground by ViBe, neighbours in a square region R were considered centered in this pixel. The candidate pixels to be classified as shadows are those whose value of correlation is high, considering this region of neighbours:

$$C(i,j) = \sum_{n=-N}^{N} \sum_{m=-N}^{N} (I(i+n, j+m) * B(i+n, j+m)) \tag{1}$$

where I(x, y) and B(x, y) is the intensity value of the pixel in images I and B, respectively; and the size of R is $(2N+1)^2$.

The correlation gets normalized as:

$$NCC(i,j) = \frac{C(i,j)}{\sqrt{mI(i,j)} * \sqrt{(mB(i,j)}} \tag{2}$$

where mI(i, j) and mB(i, j) are the second central moment within the same region R centered in the pixel (i,j), in image I y B, respectively.

The reference value to detect shadows with NCC (Eq. 2) is a high value defined within the [0.95–0.98] range [13], being used as a previous step to detect possible candidates.

Then, pixels are rectified using statistics in region R considering the analysis of the relationship between the values of intensity corresponding to images I and N for each pixel (i, j)

$$\alpha(i,j) = \frac{I(i,j)}{B(i,j)} \tag{3}$$

The areas of shadow have to be adjusted to a definite range of values, where α in Eq. 3 is usually defined in [0.4–1] range, because of the darkness of the shadow over the background model [13]. Furthermore, the standard deviation of the α relationship calculated in region R of neighbours must have a low value since the area of shadow is a homogeneous region. It is advisable to consider a standard deviation lower than 0.05 in region R bearing a size of 5×5 pixels, and N equal to 2. The problem of the original method [13] is that it generates false negatives whenever the value of the intensity of the object coincides with the values of the shadow.

Figure 4 depicts a frame of the video *Pedestrians* [14] and another frame of a video of urban surveillance (right). The first row in Fig. 4 highlights the result of the background subtraction with ViBE. It can also be observed in the case of the pedestrian that the shadow (false positives) is not connected to the body; and in the case of the vehicle, the shadow is projected in the lower part, deforming the object.

In the second row in Fig. 4, parameters recommended for NCC (Eq. 2) larger than 0.95 and for α within the range [0.4–1] were used. It can be observed that shadow detection may not always be right, as happens on the legs of the pedestrian, and also, on the glasses of the vehicle, generating false negatives. The method is not working properly, mainly in the scenes where there is light reflex, or when the background resembles the shadow. In these cases, different values of α and standard deviation were considered, depending on the noise in the video signal, and in cases where the shadow turns darker.

Fig. 4. Background subtraction and shadow detection. Original image or Object contour with detected shadow (left) and resulting mask (right). ViBE algorithm (top). Shadow detection algorithm with NCC (center) and Shadow detection with the proposed algorithm (below)

4.2 Shadow Elimination Based on Object Information

The purpose of this project was not to apply previous rules for NCC in potential areas of the true object, Firstly, shadow classification is performed at the level of a rectangular region or detected blob (components connected to the foreground mask detected by ViBE) and not of each isolated pixel as it happens in the original version of the method [13]. Therefore, the suggestion is to utilize the spacial information within this region in image I. When working with videos of people or vehicles from a video with a fixed camera, it was considered to include further knowledge of some of their properties such as orientation, shape or size.

If the case is to enclose people or vehicles within an ellipse using the mass center of the foreground mask, it is highly likely that the projection of the shadow exceeds the limits of the ellipse as it is the case of the vehicle in Fig. 4. Conversely, the projection of the shadow may remain separated in a different ellipse as it is the case of the pedestrian.

Directional distribution is calculated in each blob to find the ellipse that best adjusts to the distribution of pixels foreground. As a result of this method, an ellipse is defined through its major axis, the minor axis and the orientation of the major axis as regards the horizontal axis [15]. Furthermore, these values had to be adapted to cover all of the objects, both people and vehicles; and at the same time, not to exceed the limits of the blob. Heuristically, a factor of 1.8 became the right one for the major axis and the 1.6 for the minor axis.

As a first approach not to invade in the sequence the areas of the real object, it was considered that the height and width of the human body keep a certain proportion and that a person walks erected. Thus, the rule of decision to determine that there is a person in a blob is by the size of the area, the relationship between the major and minor diameter of the ellipse being greater or equal to 3, and finally, that the orientation of the major axis is kept close to 90°. Figure 4 above depicts the major diagonal of the ellipse being almost vertical for a person and almost horizontal for the major diagonal of the ellipse that encloses the shadow. To identify vehicles, orientation may vary; however, the size of the blob for this case is larger than that which contains people, bicycles or motorcycles.

Last row in Fig. 4 shows that the results of the proposed algorithms improve as regards the original algorithm of ViBE+ (Fig. 4, above) and also, of ViBE+ [13] (Fig. 4, center). The pedestrian has no shadow over their legs. Moreover, no shadows are considered on the windows of the vehicles, false negatives. Whenever a shadow is detected outside the ellipse, the algorithm follows over the neighbouring pixels maintaining its continuity even if it overpasses the pixels the borders of the ellipses.

5 Results

The proposed algorithm was tested with the real videos *Pedestrians* and *Highway* of data base [14,16], which were taken with static cameras used in videosurveillance. These videos have images groundtrue to calcule the rate of accuracy and compare results. Hence, the library of classic algorithms of background subtraction allows comparing with the recent methods and contributions in the area. As for the present study case, the obtained results are analyzed with some other traditional stochastic methods like VIBE+ and Gaussian Mixture Model (GMM) [3].

Each method was analyzed through two wellknown evaluation metrics Precision, Recall as:

$$Precision = \frac{TP}{(TP + FP)} \tag{4}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{5}$$

where FN the number of false negatives, FP the number of false positives, TP the total number of positives and TN the total number of negatives. Getting in this case a better result metrics when approach to 1. Also, an indicator over errors on shadows as,

$$TShadow = \frac{nbErrorShadows}{FP} \tag{6}$$

where nbErrorShadows the total of false positives produced in the area of shadows according to the groundtrue, and FP the number of false positives.

Table 1 depicts the values of the equations detailed above for approximately 700 frames of *Pedestrians* video and, in Table 2 the values for 1300 frames of *Highway* video.

Table 1. Values of indicators for each method with video Pedestrians

Methods	Recall	Precision	Tshadow
GMM [17]	0.98	0.93	0.22
ViBE+ [10]	0.93	0.96	0.50
ViBE+NCC [13]	0.83	0.99	0.11
Proposed ViBE+NCC+Information	0.91	0.99	0.07

Table 2. Values of indicators for each method with video Highway

Methods	Recall	Precision	Tshadow
GMM [17]	0.89	0.91	0.81
ViBE+ [10]	0.84	0.92	0.95
ViBE+NCC [13]	0.58	0.94	0.47
Proposed ViBE+NCC+Information	0.85	0.94	0.55

It can be observed for both videos *Pedestrians* and *Highway*, that by using the method ViBE+ o GMM, the percentage TShadow is much higher, close to the double, than the proposed method. Method ViBE+NCC [13] increases the amount of false negatives getting the Recall metric to considerably diminish in regard to ViBE+ (for *Pedestrians* from 0.93 to 0.83 and for *Highway* from 0.84 to 0.58). Furthermore, both videos with the proposed method, it may be noticed that there is an increase in the Precision with regard to ViBE+, indicating a reduction in the amount of false positives. Finally, with the proposed method ViBE+NCC+information, the rate Recall and Precision is comparable to that of the methods ViBE+ and GMM; however, the percentage of those wrongly classified in the area of shadows with TShadows is notoriously smaller.

6 Conclusions

The present research work shows the preliminary results of the method of detection and separation of the shadow, based on contextual information, which are promising. The videos have been processed and visualized their classification in real time since the algorithm has computational low cost. It was also possible to diminish the error in the classification in regard to the application of traditional NCC, calculating the orientation and probable location of the object and determining either people or vehicle.

Future works will attempt to incorporate the automatic adaptation of the thresholds for α range, considering the variation of the intensity of the shadow when processing the video, as the algorithm has difficulty in finding areas of darker shadows. Another challenge aims to achieve that the parameters of the algorithm automatically adapt to different times of the day and, and comparing quantitatively with some other reference methods.

References

1. Dominguez, L., Perez, A., Rubiales, A., Damato, J., Barbuzza, R.: Herramientas para la detección y seguimiento de personas a partir de cámaras de seguridad. In: Proceedings del XXII Congreso Argentino de Ciencias de la Computación, pp. 251–260 (2016)
2. Kruegle, H.: CCTV Surveillance: Video Practices and Technology. Butterworth-Heinemann, Newton, MA, USA (2014)
3. Legua, C.: Seguimiento automático de objetos en sistemas con múltiples cámaras (2013)
4. Shaikh, S.H., Saeed, K., Chaki, N.: Moving Object Detection Using Background Subtraction, pp. 15–23. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07386-6
5. Azab, M., Shedeed, H., Hussein, A.: A new technique for background modeling and subtraction for motion detection in real-time videos. In: IEEE International Conference on Image Processing, pp. 3453–3456 (2010)
6. Hadi, R., Sulong, G., George, L.: Vehicle detection and tracking techniques: a concise review. Sig. Image Process. Int. J. **5**, 1–12 (2014)
7. Fredembach, C., Finlayson, G.: Simple shadow removal. In: Proceedings of the 18th International Conference on Pattern Recognition (ICPRA 2006), pp. 832–835 (2006)
8. Qin, R., Liao, S., Lei, Z., Li, S.Z.: Moving cast shadow removal based on local descriptors. In: 20th International Conference on Pattern Recognition, pp. 1377–1380 (2010)
9. Chen, F., Zhu, B., Jing, W., Yuan, L.: Removal shadow with background subtraction model vibe algorithm. In: 2013 2nd International Symposium on Instrumentation and Measurement, Sensor Network and Automation (IMSNA), pp. 264–269 (2013)
10. Barnich, O., Van Droogenbroeck, M.: Vibe: a universal background subtraction algorithm for video sequences. IEEE Trans. Image Process. **20**(6), 1709–1724 (2011)

11. Gervasoni, L., Damato, J., Barbuzza, R., Vénere, M.: Un método eficiente para la sustracción de fondo en videos usando gpu. Enief 2014, Revista Mecánica Computacional, pp. 1721–1731 (2014)
12. Barbuzza, R., Damato, J., Rubiales, A., Dominguez, L., Perez, A., Venere, M.: Un método para la sustracción de fondo en videos inestables. Enief 2014, Revista Mecánica Computacional, pp. 3409–3417 (2016)
13. Jacques, J.C.S., Jung, C.R., Musse, S.R.: Background subtraction and shadow detection in grayscale video sequences. In: XVIII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2005), pp. 189–196 (2005)
14. Goyette, N., Jodoin, P.M., Porikli, F., Konrad, J., Ishwar, P.: Changedetection.net: a new change detection benchmark dataset. In: Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1–8 (2012)
15. Lahoz Beltra, R., Ortega, J., Fernandez Montraveta, C.: Métodos estadísticos en biología del comportamiento. Editorial Complutense (1994)
16. Sobral, A., Vacavant, A.: A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. Comput. Vis. Image Underst. **122**, 4–21 (2014)
17. Zivkovic, Z.: Improved adaptive Gaussian mixture model for background subtraction. In: 2004 Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004), vol. 2, pp. 28–31 (2004)