SHADOW DETECTION IN VIDEO SURVEILLANCE BY MAXIMIZING AGREEMENT BETWEEN INDEPENDENT DETECTORS*

Juan C. SanMiguel, José M. Martínez

Video Processing and Understanding Lab
Escuela Politécnica Superior, Universidad Autónoma de Madrid, Spain
E-mail: {JuanCarlos.SanMiguel, JoseM.Martinez}@uam.es

ABSTRACT

This paper starts from the idea of automatically choosing the appropriate thresholds for a shadow detection algorithm. It is based on the maximization of the agreement between two independent shadow detectors without training data. Firstly, this shadow detection algorithm is described and then, it is adapted to analyze video surveillance sequences. Some modifications are introduced to increase its robustness in generic surveillance scenarios and to reduce its overall computational cost (critical in some video surveillance applications). Experimental results show that the proposed modifications increase the detection reliability as compared to some previous shadow detection algorithms and performs considerably well across a variety of multiple surveillance scenarios.

Index Terms— Shadow detection, mutual information, detectors agreement, video processing, video surveillance.

1. INTRODUCTION

Nowadays, surveillance systems have more demand, specifically for its application in public areas, as airports, stations, subways, entrance to buildings and mass events. In this context, reliable detection of moving objects is the most critical requirement for the surveillance systems.

In the moving object detection process, one of the main challenges is to differentiate moving objects from their shadows. Moving cast shadows [1] are usually misclassified as part of the moving object making the following analysis stages, such as object classification or tracking, to perform inaccurate. Several shadow detection algorithms have been proposed and they can be classified by their use of chromaticity information [2][3][4], edge information [5], stereo information [6] or a combination between them [7].

Moreover, the shadow detection process usually involves a number of classifiers which are trained with labelled data. In this context, the availability of training data is a critical issue and its creation presents two basic problems: the difficulty of manual annotation (determining the accuracy of the learned models) and the amount of data used (the classifier will be very specific if it is huge or it won't be optimal if it is small). Some approaches have been proposed to avoid the need of training data and therefore the problems mentioned above. The approach commonly used is based on maximizing the agreement between independent detectors applied on the same data. In [8], the authors adaptively compute thresholds for foreground detection maximizing the mutual information between foreground maps calculated for visual and thermal infrared images. The same authors apply the previous scheme for the shadow detection task in single images in [4]. Similarly, in [3] an HSV-based shadow detection algorithm is proposed that dynamically estimates its parameters using some pre-defined relations between the object and shadow pixels that are independent of the scene type.

In this paper we extend the work presented in [4]. Firstly, the shadow detection algorithm for single images is described and adapted to process video surveillance sequences. Then, some modifications are included to increase the robustness and reduce the computational cost of the adapted shadow detection algorithm. Finally, the proposed algorithm is tested with indoor/outdoor video surveillance sequences to analyze the modifications added.

The paper is structured as follows: section 2 briefly describes the idea presented in [4], section 3 describes the adaptation to analyze video sequences and section 4 the modifications done to increase the overall performance. In section 5, experimental results and comparative ones are shown, while section 6 closes the paper with some conclusions and future work.

*Work supported by the Spanish Government (TEC2007- 65400 SemanticVideo), by Cátedra Infoglobal-UAM for "Nuevas Tecnologías de video aplicadas a la seguridad", by the Spanish Administration agency CDTI (CENIT-VISION 2007-1007), by the Comunidad de Madrid (S-050/TIC-0223 - ProMultiDis), by the Consejería de Educación of the Comunidad de Madrid and by the European Social Fund.

2. BASE SHADOW DETECTION ALGORITHM

The shadow detection algorithm described in [4] (from now on base algorithm) is based on the application of two independent shadow detectors with different configurations (thresholds) and chooses the one that maximizes the agreement between the applied detectors. Basically, the idea is to train each detector with the other (finding the maximum agreement). This process is iteratively repeated until the parameter configuration has no change (optimum configuration).

In this algorithm, there are three relevant aspects: the shadow detectors applied, the agreement measure and the optimization algorithm.

As independent shadow detectors, they propose to use the bounded decrease in brightness (using two thresholds) and the bounded decrease in saturation (using two thresholds). These two decreases are computed as the relative change between each pixel in the current and background image (calculated as the median background image) using the HSV colour space. As the output of the shadow detectors is a binary image, the agreement measure is computed between two binary signals using the Kendall's τ or the mutual information [4]. These two measures compute the correlation between the two binary signals (masks) using a 4-value co-occurrence histogram of the pixels values. The optimization algorithm is used to search the maximum of the agreement measure. The overall optimization process is based on the interaction between two optimization stages until there is no change. Each stage tries to choose the optimum thresholds of one detector considering the output image of the other detector constant. The authors propose to use a dynamic programming-based solution that reduces the number of iterations to find the desired value of thresholds that maximize the agreement.

3. VIDEO PROCESSING SYSTEM OVERVIEW

The shadow detection algorithm described above is used in single images without any pre/post-processing stages.

In this paper we propose to introduce the algorithm in a video surveillance system. Firstly, as a pre-processing stage, we propose to introduce a background subtraction stage (using a standard GMM algorithm [9]) that reduces the amount of data analyzed in the shadow detection stage. Then, we propose a noise removal stage based on mathematical morphology as a post-processing stage of the foreground and shadow detection stages. Specifically, the operation used is called "Opening by reconstruction of erosion" [10] and it preserves the underlying shape of the object associated to the shadows. Furthermore, the results of each optimization stage (optimum thresholds) are used as starting point in the analysis of the following frame. Additionally, a buffer is introduced in the optimization stage to speed-up the process.

4. MODIFICATION DESCRIPTION

In this section we describe some modifications introduced to increase the algorithm performance in many situations and reduce its computational cost.

4.1. Addition of a new shadow detector

The algorithm described in [4] presents some false positives due to the detection of isolated pixels as shadows.

For this reason, we propose to introduce a new shadow detector based on the hypothesis that the intensity reduction inside a shadow region is similar. It is based in the colour constancy property between pixels of the same region and it is described in [7]. This detector is introduced to discard the wrong detection of shadow pixels in the base algorithm. Additionally, the base algorithm already computes the intensity reduction ratio (see section 2) and the calculation of this new detector adds low computational cost to the whole shadow detection algorithm.

4.2. Agreement measure

In [4], the agreement measure used (Kendall's τ) is described between two binary sources. In the proposed algorithm, there are three independent binary sources (with the addition of the shadow detector described in 4.1) and the calculation of the Kendall's τ (extended to three binary signals) significantly increases the computational cost.

In [11], the authors propose to calculate binary similarity measures between different bit planes of an image. The basic idea is to compute for each pixel the similarity of each bit plane with the other bit planes and to integrate all the computed similarities. As the calculation of these measures requires a lot of operations, their calculation is reasonable when we have a large number of binary sources.

In this paper we propose to use a simple and well known measure of similarity, the correlation between two signals. This operation is very efficient in computational terms and it can be easily extended to three binary signals (computing it in pairs).

4.3. Optimization stage

In the base algorithm [4], the described optimization stage is based on the interaction and optimization between two sources. It can be extended to deal with three binary sources but the computational cost is highly increased. Additionally, the computational cost of the optimization stage used can be reduced by selecting an initial value close to the optimum thus reducing the value search range as in video surveillance sequences values are similar in consecutive frames.

In this paper, we propose to use a gradient ascent algorithm [12] to choose the optimum thresholds between the three detectors (using 4 thresholds that correspond to maximum and minimum decrease in brightness, maximum difference in saturation and the maximum intensity reduction ratio). Additionally, we have divided the optimization process in two stages to speed-up it. The first stage uses a coarse step to choose values close to the optimum and the second stage uses a fine step to choose the optimum values.

4.4. Temporal filtering

Experiments performed on the base algorithm showed that sometimes the results (the optimized thresholds) are not determined correctly. This is due to the shadow strength, shadow similarity with the background, shadow pixels percentage with respect to the foreground pixels and other issues. These failures are difficult to detect and the use of the agreement measure (or the quality of the agreement [4]) does not provide enough information to detect them.

In this paper, we propose to use a temporal filtering stage to reduce the effect of these wrong optimum values determined using a Gaussian filter for smoothing each optimized threshold

4.5. Fusion scheme

The addition of a third detector in the shadow detection process allows the combination of the results of the shadow detectors using different schemes.

In this paper, we propose to use two simple fusion schemes for considering a pixel as a shadow. One is based on the agreement between the three detectors and the other is based on the agreement between two (at least) of them.

5. EXPERIMENTAL RESULTS

In this section, experimental results of the proposed algorithm are presented.Experiments were carried out on selected sequences from the i-LIDS dataset for AVSS2007 (available at http://www.avss2007.org), the PETS2006 dataset (available at http://www.pets2006.net/), the ATON dataset (available at http://cvrr.ucsd.edu/aton/shadow/). The system has been implemented in Matlab. Tests were executed on a Pentium IV with a CPU frequency of 2.8 GHz and 1GB RAM.

Dataset	Sequence	Type	Length	Shadow Strength	Noise Level	Shadow Size
ATON	Int. Room	Indoor	900	Low	High	Medium
	Campus	Outdoor	1179	Low	High	Large
AVSS 2007	PV_Easy	Outdoor	1200	Medium	Low	Medium
	AB_Easy	Indoor	2000	Low	Low	Small
PETS 2006	S1_T1_C3	Indoor	3020	Medium	Low	Large
	S4_T5_A3	Indoor	2195	Medium	Medium	Large

Table 1: Description of properties of each test sequence

In order to evaluate the performance of the proposed algorithm, we have determined the properties of the selected sequences. A description of the selected test sequences and its properties is shown in Table 1. Following other authors [1][2][3][5], we have decided to use the *shadow detection accuracy* η and the *shadow discrimination accuracy* ξ as the performance measures. They are defined as follows:

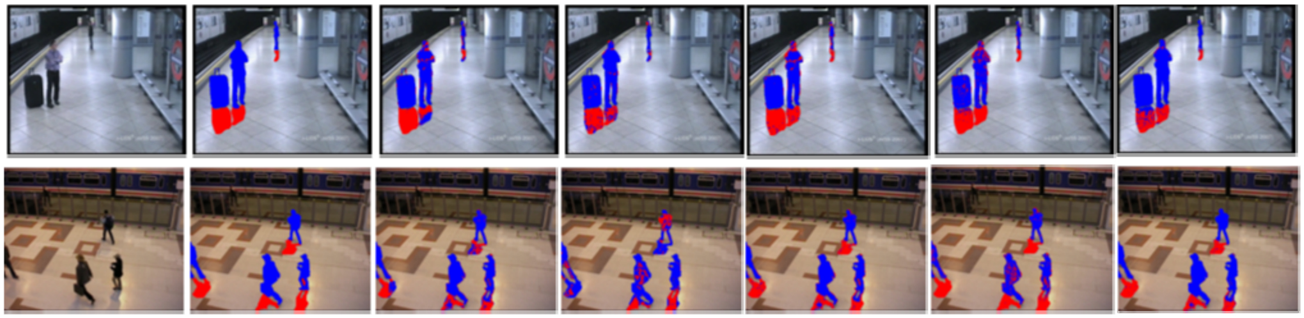
$$\eta = \frac{TP_S}{TP_S + FN_S} \quad (1) \qquad \xi = \frac{TP_F}{TP_F + FN_F} \quad (2)$$

where TP indicates *True Positives*, FN indicates False *Negatives* and the subscript S/F indicates Shadow or Foreground.

To evaluate and compare the performance of the proposed algorithm, we provide labeled data to it, therefore decoupling the errors coming from the background subtraction stage from the ones inherent to the proposed algorithm. Firstly, a background image of each test sequence is calculated using the system proposed in section 3 and then a foreground mask is calculated using the foreground and shadow ground-truth data. This ground-truth has been prepared on representative frames of the different situations of interest for each video sequence (different shadow size/strength, low percentage of shadow points…). The experimental results obtained using the two fusion schemes proposed are compared with respect to the standard HSV shadow detection algorithm [2] using the fixed thresholds proposed in [2], the base algorithm [4] (we want to thank the authors for the code provided for running these set of tests) and its adaptation. They are summarized

	ATON				AVSS 2007				PETS 2006				Average framerate (320x240)
	Campus		Int. Room		PV_Easy		AB_Easy		S1_T1_C3		S4_T5_A3		
	η (%)	ξ (%)	η (%)	ξ (%)	η (%)	ξ (%)	η (%)	ξ (%)	η (%)	ξ (%)	η (%)	ξ (%)	
Standard HSV algorithm [2]	25.3	69.6	35.1	85.5	68.6	88.2	30.2	87.5	37.6	89.4	29.3	94.8	12 fps
Base algorithm [4]	17.7	92.6	42.6	54.8	60.5	44.7	73.0	76.9	39.2	66.4	56.0	59.1	1.05 fps
Base algorithm [4] adapted	42.1	68.7	57.5	86.9	67.3	52.8	89.8	86.29	85.6	83.6	84.3	93.0	1.55 fps
Proposed algorithm (agreement between two)	69.6	79.6	80.9	83.1	70.5	77.6	95.8	86.0	98.0	74.0	94.0	91.1	2.20 fps
Proposed algorithm (agreement between three)	43.9	98.1	60.8	94.9	42.9	98.5	81.2	98.5	88.0	90.8	78.8	98.8	2.20 fps

Table 2: Comparative Results (in percentage)



Original Ground-Truth Standard HSV Base algorithm Base Alg. Adapted Proposed (A.2) Proposed (A.3)

Figure 1: Comparative Results for frame 2457 of AB_Easy sequence and frame 1030 of S4_T5_A3 sequence

in Table 2. Additional results are available at http://www-vpu.ii.uam.es/publications/ICIP09ShadowDetection .

As shown in Table 2, the proposed algorithm improves the detection results of the standard HSV [2] and the base algorithm [4] adapted to process video sequences. It can be observed that [2] presents a low detection performance showing the dependency of the parameters with the sequence being analyzed. On the other hand, the base algorithm automatically chooses the optimum parameters improving the results of [2] in all the test sequences. However, its *shadow detection accuracy* is quite low in sequences with high noise level (*Campus* and *Intelligent Room)*. The adaptation proposed for the base algorithm increases the performance in all the test sequences mainly due to the reduction of the amount of data to analyze. The proposed algorithm improves the overall results of the base algorithm adapted in all the test sequences. The addition of a new shadow detector combined with the fusion schemes allows to select which performance measure is improved, selecting the agreement between the three detectors to improve the foreground detection results or the agreement between two of the detectors to improve the shadow detection results. Finally, we have evaluated the computational cost of the compared algorithms (see Table 2). It can be observed that the standard HSV algorithm [2] presents the lower computational cost as the parameters are fixed and there is no optimization process. The proposed algorithm is faster than the base algorithm and its adaptation due to the modifications included in the agreement measure and in the optimization stage.

6. CONCLUSIONS

In this paper a new approach for robust shadow detection in video surveillance is presented. It dynamically estimates the optimum thresholds for each video sequence without training data and it is based on the maximization of the agreement between the detectors applied as described in [4]. Some improvements are presented, which increase the robustness of the algorithm and reduce the computational cost. Experimental results show that the proposed scheme is significantly more efficient and stable than the standard HSV [2] and the base algorithm [4].

Future work includes an extension of the proposed algorithm including a complex scheme that integrates temporal and spatial information about shadow pixels detected. Additionally, like the authors of [4] suggest, we want to use the described algorithm to dynamically select the optimum color-space from the available ones.

7. REFERENCES

[1] Prati, A.; Mikic, I.; Trivedi, M.M.; Cucchiara, R., "*Detecting moving shadows: algorithms and evaluation*", IEEE Trans. on Pattern Analysis and Machine Intelligence, 25(7):918-923, 2003.

[2] Cucchiara, R.; Grana,C.; Piccardi,M.; Prati,A. "*Detecting moving objects, ghosts and shadows in video streams*", IEEE Trans. on Pattern Analysis and Machine Intelligence, 25(10):1337-1342, 2003.

[3] Tattersall, S. and Dawson-Howe, K., "*Adaptive Shadow Identification through Automatic Parameter Estimation in Video Sequences*," Proc. of Irish Machine Vision and Image Processing Conf., pp. 57-64, 2003.

[4] Conaire, C.; O'Connor, N.; Smeaton, A., "*Detector adaptation by maximising agreement between independent data sources*," Proc of IEEE Conf. on Computer Vision and Pattern Recognition, pp.1-6, 2007.

[5] Xu D.; Liu, J.; Li, X.; Liu,Z.; Tang X., "*Insignificant shadow detection for video segmentation*," IEEE Trans. on Circuits and Systems for Video Technology, 15(8):1058-1064, 2005.

[6] Haritaoglu, I. "*A real time system for detection and tracking of people and recognizing their activities*", PhD thesis, University of Maryland at College Park, 1998.

[7] Yang, M.-T.; Lo, K.-H.; Chiang, C.-C.; Tai, W.-K., "*Moving cast shadow detection by exploiting multiple cues*," IET Image Processing, 2(2):95-104, 2008.

[8] Conaire, C.; O'Connor, N.; Cooke E; Smeaton, A., "*Detection Thresholding Using Mutual Information*", Proc of Int. Conf. on Computer Vision Theory and App., pp 408-415, 2006

[9] Stauffer, C.; Grimson, W., "*Learning Patterns of Activity Using Real-Time Tracking*", IEEE Trans. on Pattern Analysis and Machine Intelligence, 22(8): 747-757, 2000

[10] Salembier, P. and Ruiz, J., "*On filters by reconstruction for size and motion simplification*", Proc. of Int. Symposium in Mathematical Morphology, pp. 425-434, 2002

[11] Bayram S.; Avcıbaş, I.; Sankur, B.; Memon,N. "*Image Manipulation Detection with Binary Similarity Measures*," Proc of European Signal Processing Conf., 752-755, 2005.

[12] Snyman, J.A., "*Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical New Gradient-Based Algorithms*". Springer Publishing, 2005.