Shadow Detection Using 2D Cepstrum

B. Ugur Toreyin*, A. Enis Cetin
Bilkent University, Bilkent 06800, Ankara, Turkey
{ugur, cetin}@ee.bil.kent.edu.tr

ABSTRACT

Shadows constitute a problem in many moving object detection and tracking algorithms in video. Usually, moving shadow regions lead to larger regions for detected objects. Shadow pixels have almost the same chromaticity as the original background pixels but they only have lower brightness values. Shadow regions usually retain the underlying texture, surface pattern, and color value. Therefore, a shadow pixel can be represented as a.x where x is the actual background color vector in 3-D RGB color space and a is a positive real number less than 1. In this paper, a shadow detection method based on two-dimensional (2-D) cepstrum is proposed.

Keywords: Shadow detection in video, 2-D Cepstrum

1. INTRODUCTION

Moving shadow and shaded regions are major sources of false alarms for most of the moving object detection and tracking algorithms in video [1-3]. These algorithms basically depend on tracking the change in the intensity values of pixels over time. Similar to regular moving objects, moving shadow regions have variations in intensity values which results in false alarms. In addition to these false alarms, moving shadow areas within the viewing range of cameras yield larger regions for detected video objects.

Recently, there is an increased interest in camera-based forest fire detection systems monitoring forestry all around the world [4-9]. These systems utilize computer vision algorithms to detect forest fires. It is reported in a recent study that, shadows of slow moving clouds are major source of false alarms for these systems, as well [9]. In order to reduce false alarm rate, some sort of shadow detection and elimination method should be incorporated to these systems.

It is well-known in the computer vision literature that, shadow and shaded areas retain underlying texture, surface pattern, color and edges in images [10, 11]. Apparently, pixels in shadow regions have almost the same chromaticity value as the original background pixels. However, they have lower brightness values compared to the original background pixels. In this paper, a cepstrum analysis based shadow detection method is proposed. Cepstrum based signal processing techniques have been widely used especially in speech processing and recognition field [12-14]. Cepstrum analysis is a non-linear technique involving natural logarithm operation.

Proposed method comprises of two steps: 1. Hybrid background subtraction based moving object detection and 2. Cepstrum analysis based shadow detection. Moving object detection method in the first step of the proposed method determines the candidate regions for further analysis. The second step involves a non-linear method based on cepstrum analysis of the pixels in the candidate regions.

This paper is organized as follows, in the next section, the two steps of the proposed moving shadow detection method is presented. In Section 3, experimental results obtained with forest monitoring videos are discussed. Conclusions are drawn in the last section.

2. MOVING SHADOW DETECTION

The proposed method for moving shadow detection in video is composed of two sub-algorithms, namely: 1. Hybrid background subtraction based moving object detection and 2. Cepstrum analysis based shadow detection. In the first phase of the moving shadow detection method, moving objects are detected by a hybrid background subtraction method. A cepstrum based analysis is carried out for moving pixels. This analysis yields regions with shadows. These methods are discussed in the following sub-sections.

2.1 Hybrid Background Subtraction Method

Moving object detection method described in [VSAM] is used for motion detection in video. Background subtraction is commonly used for segmenting out moving objects in a scene for surveillance applications. There are several methods in the literature [1-3] for moving object detection in video. The background estimation algorithm described in [1] uses a simple IIR filter applied to each pixel independently to update the background and use adaptively updated thresholds to classify pixels into foreground and background.

Stationary pixels in the video are the pixels of the background scene because the background can be defined as temporally stationary part of the video. If the scene is observed for some time, then pixels forming the entire background scene can be estimated because moving regions and objects occupy only some parts of the scene in a typical image of a video. A simple approach to estimate the background is to average observed image frames of the video. Since moving objects and regions occupy only a part of the image, they conceal a part of the background scene and their effect is canceled over time by averaging. In Video Surveillance and Monitoring (VSAM) Project at Carnegie Mellon University [1] a recursive background estimation method was developed from the actual image data using "l1-norm" based calculations.

Let $I(x, n)$ represent the intensity value of the pixel at location x in the n-th video frame I. Estimated background intensity value, $B(x, n + 1)$, at the same pixel position is calculated as follows:

$$B(x, n+1) = \begin{cases} aB(x,n) + (1-a)I(x,n) & \text{if } x \text{ is stationary} \\ B(x,n) & \text{if } x \text{ is a moving pixel} \end{cases}$$

(1)

where $B(x, n)$ is the previous estimate of the background intensity value at the same pixel position. The update parameter a is a positive real number close to one. Initially, $B(x, 0)$ is set to the first image frame $I(x, 0)$. A pixel positioned at x is assumed to be moving if:

$$|I(x, n) - I(x, n-1)| > T(x, n)$$

(2)

where $I(x, n-1)$ is the intensity value of the pixel at location x in the (n-1)-th video frame I and $T(x, n)$ is a recursively updated threshold at each frame n, describing a statistically significant intensity change at pixel position x:

$$T(x, n+1) = \begin{cases} aT(x,n) + (1-a)(c|I(x,n) - B(x,n)|) & \text{if } x \text{ is stationary} \\ T(x,n) & \text{if } x \text{ is a moving pixel} \end{cases} \qquad (3)$$

where c is a real number greater than one and the update parameter a is a positive number close to one. Initial threshold values are set to a pre-determined non-zero value.

Both the background image $B(x, n)$ and the threshold image $T(x, n)$ are statistical blue prints of the pixel intensities observed from the sequence of images $\{I(x, k)\}$ for $k < n$. The background image $B(x, n)$ is analogous to a local temporal average of intensity values, and $T(x, n)$ is analogous to c times the local temporal standard deviation of intensity in "l1-norm" [1].

As it can be seen from Eq. (3), the higher the parameter c, higher the threshold or lower the sensitivity of detection scheme. It is assumed that regions significantly different from the background are moving regions. Estimated background image is subtracted from the current image to detect moving regions which corresponds to the set of pixels satisfying:

$$|I(x, n) - B(x, n)| > T(x, n) \qquad (4)$$

are determined. These pixels are grouped into connected regions (blobs) and labeled by using a two-level connected component labeling algorithm [15]. The output of the first step of the algorithm is a binary pixel map $Blobs(x,n)$ that indicates whether or not the pixel at location x in n-th frame is moving.

Other more sophisticated methods, including the ones developed by Bagci et al. [2] and Stauffer and Grimson [3], can also be used for moving pixel estimation. In our application, accurate detection of moving regions is not as critical as in other object tracking and estimation problems; we are mainly concerned with real-time detection of moving shadows as part of shadow elimination step of a computer vision based forest fire detection system.

2.2 Cepstrum Analysis

Cepstrum analysis is applied for candidate moving shadow pixels detected by the first step of the algorithm. It involves natural logarithm and is a non-linear technique commonly used in speech processing applications [12-14]. The complex cepstrum of a signal x is the inverse Fourier transform of the complex natural logarithm of the Fourier transform of x. Let $x[n]$ be a discrete signal, as in our case, then, the complex cepstrum $x^\wedge[n]$ of $x[n]$ is defined as:

$$x^\wedge[n] = F^{-1}\{ln\{F\{x[n]\}\}\} \qquad (5)$$

where $F\{.\}$ is the discrete-time Fourier transform, $ln\{.\}$ is the natural logarithm and $F^{-1}\{.\}$ is the inverse Fourier transform operators.

Let a pixel x inside a moving shadow region have the following background, $x\_bg = (R\_bg, G\_bg, B\_bg)$, and image values $x\_im = (R\_im, G\_im, B\_im)$ values. As discussed in [11], ideally shadow retains color and texture in images. According to this assumption, the following equation should hold for background and current image pixel values of a shadow region:

$$x\_im = d\,x\_bg \qquad (6)$$

where d is a positive real number less than one. Assuming Eq. (6) holds, second and third 2D cepstral coefficients $x\_im^[2]$ and $x\_im^[3]$ corresponding to x_im and their counterparts $x\_bg^[2]$ and $x\_bg^[3]$ corresponding to x_bg should be the same. This can ideally be represented as:

$$x\_im^[2] = x\_bg^[2] \qquad (7)$$

and

$$x\_im^[3] = x\_bg^[3] \qquad (8)$$

First cepstral coefficients $x\_im^[1]$ and $x\_bg^[1]$ of x_im and x_bg are different due to the additional effect of the natural logarithm of the coefficient d in Eq. (6).

Based on this assumption, second and third complex cepstral coefficients of moving pixels in the current and background image frames are calculated and compared with each other. If the values are close to each other, then these pixels are determined as moving shadow pixels within the viewing range of the camera.

3. EXPERIMENTAL RESULTS

The proposed two-step method for moving shadow detection is tested with 17 clips which are recorded from a visible-range camera mounted on top of a forest watch tower near Antalya region in Turkey. The test set consists of a total of 4250 image frames. In the test set, there are two clips with genuine forest fire smoke and the rest of them are regular forest recordings with moving cloud shadows over forest foliage.

Shadows of moving clouds over forest foliage are the main source of false alarms in computer vision based forest fire detection system described in [4]. On the contrary, static shadow regions do not trigger any alarms. Therefore, the shadow detection method described in this paper mainly deals with the elimination of moving shadow regions within the viewing range of a camera. Sample snapshots of background and current image frames corresponding to one of the clips in the test set are shown in Figs. 1 and 2 with moving and static shadows due to clouds.

Fig. 1. Background image at frame number 100 in one of the test videos.

Fig. 2. Current image frame at frame number 100 in the same video as in Fig.1

Static shadow regions in Figs. 1 and 2 are the shaded areas that occupy the same image regions in these figures. The proposed method aims at detecting moving cloud regions. The detected shadow regions for this case are depicted in Fig.3. Moving shadow regions are marked with pale red color and encapsulated with red bounding boxes.
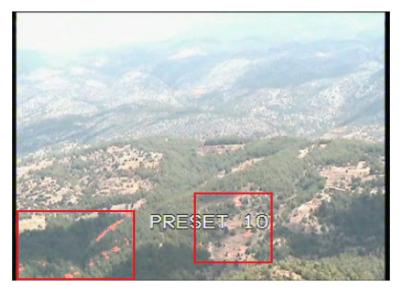
Fig. 3. Moving shadow regions are marked with pale red color and encapsulated with their minimum bounding rectangles.
The proposed method successfully detects moving shadow regions ignoring the stationary shaded areas.

No false positive alarms are issued for the genuine forest fire clips in the test set. This is extremely important for the performance of the forest fire detection method in [4]. If smoke regions due to genuine forest fires would have triggered an alarm in the proposed moving shadow detection method, then this would cause the forest fire detection system to ignore the genuine fires which may have devastating consequences.

4. CONCLUSIONS

A 2-D cepstrum analysis based method for detection of moving shadow regions in visible-range video is proposed in this paper. The method comprises a two step approach with moving object detection phase followed by further cepstral analysis step. Shadow retains underlying texture and color in moving regions. Color consistency over time is analyzed by the proposed complex cepstrum based method. Cepstral coefficients of moving regions corresponding to background and current image frames are compared to detect shaded areas within the viewing range of camera. The proposed method is tested with clips recorded from camera mounted on top of a forest watch tower monitoring a forest region. Moving cloud shadows over tree foliage are successfully detected with the proposed method.

REFERENCES

[1] Collins, R., Lipton, A. and Kanade, T., "A system for video surveillance and monitoring," 8-th International Topical Meeting on Robotics and Remote Systems. American Nuclear Society, (1999).
[2] Bagci, M., Yardimci Y. and Cetin, A., "Moving object detection using adaptive subband decomposition and fractional lower order statistics in video sequences," Signal Processing, 1941-1947, (2002).
[3] Stauffer, C. and Grimson, W. "Adaptive background mixture models for real time tracking," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2), (1999).
[4] Toreyin, B.U., Dedeoglu, Y. and Cetin, A.E. "Computer vision based forest fire detection and monitoring system," 4th International Wildland Fire Conference, Seville, Spain, 13-17 May, (2007).
[5] Dios, J. M. de, Arrue, B. , Ollero, A., Merino, L. and Gomez-Rodriguez, F., "Computer vision techniques for forest fire perception," Image and Vision Computing, 26(4), (2008).

[6] Li, J., Qi, Q., Zou, X., Peng, H., Jiang, L. and Liang, Y., "Technique for automatic forest fire surveillance using visible light image," International Geoscience and Remote Sensing Symposium, (5), 31-35, (2005).

[7] Bosch, I., Gomez, S., Vergara, L. and Moragues, J., "Infrared image processing and its application to forest fire surveillance," IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS), 283-288, (2007).

[8] Guillemant, P. and Vicente, J., "Real-time identification of smoke images by clustering motions on a fractal curve with a temporal embedding method," Optical Engineering, (40), 554-563, (2001).

[9] Toreyin, B.U. and Cetin, A.E., "Wildfire detection using LMS based active learning," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), (2009).

[10] Prati, A., Mikic, I., Trivedi, M. and Cucchiara, R., "Detecting moving shadows: Algorithms and evaluation," IEEE Transactions on Pattern Analysis and Machine Intelligence, (25), 918-923, (2003).

[11] Horprasert, T., Harwood, D. and Davis, L., "A statistical approach for real-time robust background subtraction and shadow detection," IEEE Int. Conference on Computer Vision/FRAME-RATE Workshop (CVPR/FR), (1999).

[12] Oppenheim, A. V. and Schafer, R., "From frequency to quefrency: a history of the cepstrum," IEEE Signal Processing Magazine, 9, 95-106 (2004).

[13] Furui, S. "Cepstral Analysis Technique for Automatic Speaker Verification," IEEE Trans. on ASSP, 29(2), 254-272, (1981).

[14] Childers, D.G., Skinner, P.D., and Kemerait, R.C, "The cepstrum: a guide to processing," Proceedings of the IEEE, 65(10), 1428-1443, (1977).

[15] Heijden, F., [Image Based Measurement Systems: Object Recognition and Parameter Estimation], Wiley, (1996).