Front. Electr. Electron. Eng. China 2008, 3(4): 406–410
DOI 10.1007/s11460-008-0083-6

RESEARCH ARTICLE

Baisheng CHEN

Indoor and outdoor people detection and shadow suppression by exploiting HSV color information

Abstract An adaptive background model based on maximum statistical probability and a shadow suppression scheme for indoor and outdoor people detection by exploiting hue saturation value (HSV) color information is proposed. To obtain the initial background scene, the frequency of R, G, and B component values for each pixel at the same position in the learning sequence are respectively calculated; the R, G, and B component values with the biggest ratios are incorporated to model the initial background. The background maintenance, or the so-called background re-initiation, is also proposed to adapt to scene changes such as illumination changes and scene geometry changes. Moving cast shadows generally exhibit a challenge for accurate moving target detection. Based on the observation that a shadow cast on a background region lowers its brightness but does not change its chromaticity significantly, we address this problem in the article by exploiting HSV color information. In addition, quantitative metrics is introduced to evaluate the algorithm on a benchmark suite of indoor and outdoor video sequences. The experimental results are given to show the performance of the algorithm.

Keywords background subtraction, hue saturation value (HSV) color model, shadow suppression

1 Introduction

Human motion analysis attempts to detect, track and identify people and, more generally, to understand human behaviors from image sequences involving humans. Motion detection is the key step because all consequent processes are based on its resulting regions. However, accurate motion detection exhibits a big challenge due to dynamic scene changes and disturbance.

Background subtraction techniques are mostly used for motion detection in many real-time vision surveillance applications. In these approaches, differencing between the coming frame and the background image is performed to detect foreground objects. Background subtraction provides the most complete feature data, but is extremely sensitive to dynamic scene changes due to illumination changes and extraneous events. Most researchers are now devoted to developing robust background models to prevent falseness in motion detection caused by scene changes. For example, Kalman and Brandt [1] propose an adaptive background model using Kalman filtering to adapt to the temporal variation of weather and environmental illumination. The algorithm proposed by Stauffer and Grimson [2] uses a mixture of normal distributions to model a multimodal background image sequence, and an online estimation technique is used to update the background model.

Generally, moving cast shadows exhibit a challenge for accurate moving target detection. Moving cast shadows can cause object merging, object shape distortion, and even object losses (due to the shadow cast over another object). For this reason, moving shadow detection is critical for accurate object detection in vision surveillance applications. In recent years, many algorithms have been proposed to deal with shadows. In Ref. [3], a comparative evaluation and classification of existing approaches has been elaborately presented.

In this paper, we propose an adaptive background model based on maximum statistical probability and is updated simply by background re-initiation to adapt to the scene dynamic changes. To address moving cast shadows, we propose a shadow suppression scheme utilizing the shadow features on brightness and chromaticity and implemented it in the hue saturation value (HSV) color space. The remainder of this article is organized as follows: Section 2 focuses on background initialization and maintenance. In Sect. 3, the foreground detection and shadow suppression scheme is described in depth. The

Translated from *Journal of Huaqiao University* (*Natural Science*), 2007, 28(1): 30–33 [译自: 华侨大学学报 (自然科学版)]

Baisheng CHEN (✉)
Department of Computer Science, Huaqiao University, Quanzhou 362021, China
E-mail: samchen@hqu.edu.cn

experimental results and analysis are given in Sect. 4. Finally, conclusions are drawn.

2 Adaptive background model

2.1 Background initiation

The initial background model is obtained even if there are moving foreground objects in the field of view such as walking people and moving cars. An efficient method was proposed in Ref. [4] to obtain the initial background image. In the algorithm, the frequency ratios of the intensity values for each pixel at the same position in the frames are calculated using several seconds of video (typically 10–30 s) to distinguish moving pixels from stationary ones. The intensity values for each pixel with the biggest ratios are incorporated to model the background scene. Given the learning period time t, for each pixel (x, y), let the set of different values at the position (x, y) be $\{\mu_i\}$, and set $\{\mathrm{freq}(\mu_i)\}$ be the set of corresponding frequency ratios of each value. The background initiation procedure is then illustrated as follows:

$$\forall \mu_n \in \{\mu_i\},\ \text{if}\ \mathrm{freq}(\mu_n) = \max\{\mathrm{freq}(\mu_i)\} \Rightarrow B(x,y) = \mu_n, \quad (1)$$

where $i = 1,2,\ldots,N$, N is the count of different values at the pixel (x, y) during the learning period time t.

The presented algorithm extends this idea to the RGB color space. Each component of RGB is respectively treated with the same scheme mentioned above. In our implementation, an RGB histogram is created to record the frequency of each component of the pixel. This process is illustrated in Fig. 1.

At the end of the learning stage, the biggest frequency ratio of each RGB component is estimated by using the histogram obtained above. The RGB component value with the biggest frequency ratio estimated in the histogram is assigned as the corresponding component value of the pixel in the background model. All those processes are repeated for each pixel in the background model. The output of the algorithm for several examples provided in different training sequences is shown in Fig. 2.

The idea of the algorithm is derived from the median filter but not exactly the same. It costs relatively less in terms of both memory size and time consumption than other methods based on the median filter [5]. In particular, the use of the median relies on an assumption that the background at every pixel will be visible for more than fifty percent of the time during the training sequence. For the presented algorithm, the assumption for the background visibility at each pixel turns to the biggest redundancy of the intensity values in the training sequence, which is more flexible and applicable for real events than the median filter.

2.2 Background maintenance

The difficult part of background subtraction is not the differencing itself, but the maintenance of a background model – some representation of the background and its associated statistics. Reference [6] makes a detailed discussion about problems that an ideal background maintenance system has to deal with. To adapt to scene

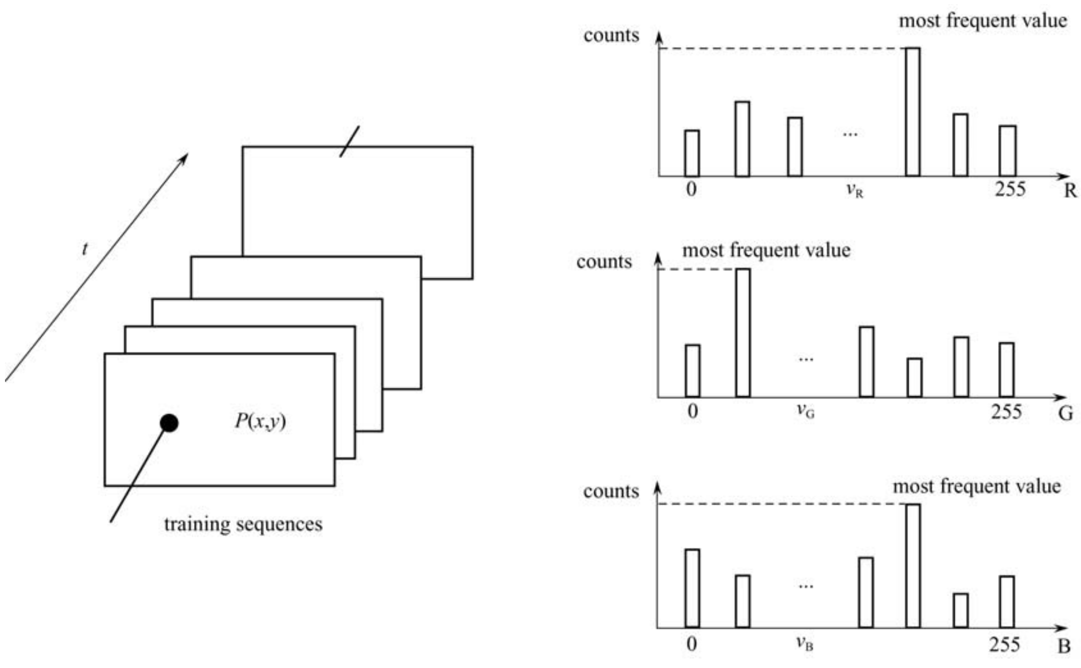

Fig. 1 RGB histograms of pixel (x, y) over learning period for background initiation

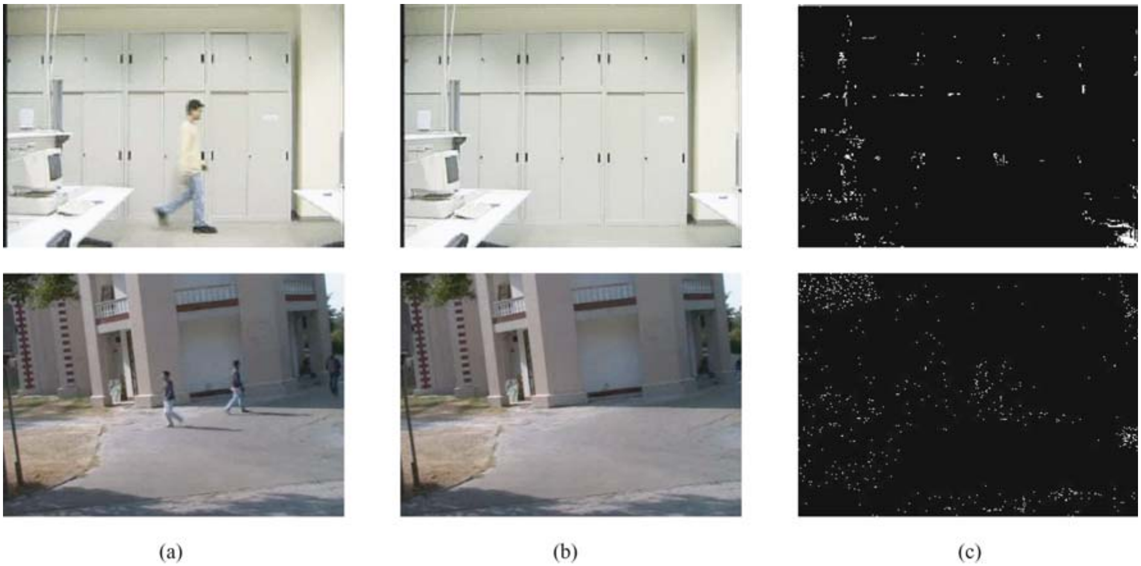

Fig. 2 Results of background initiation for indoor and outdoor sequences. (a) Training sequences; (b) background images; (c) error pixels

changes, many adaptive models are proposed for background maintenance [1,2,6], of which generally three problems may decrease the effects:

1) Intensive computation due to complexity of their algorithm;

2) Less flexibility and applicability caused by empirical parameters needed in their method which are configured to refine the behavior of the system;

3) Large errors between the computed background image and the real background scene due to methods prone to averaging the learning sequence of images.

Unlike the background maintenance methods covered above, the presented approach designs no special procedure but uses the same algorithm as the one for background initiation to update the background model periodically (100–200 frames) and adapt to illumination changes and scene geometry changes in the background scene. Thus, it is a periodical re-initialization of the background model that provides the newest background scene and produces robust results for background maintenance process. Compared with other methods, it works relatively better to adapt to scene changes in applications.

3 HSV model-based shadow suppression

3.1 Observations

In essence, shadows occur when objects partially or completely occlude direct light from a light source. As defined in Ref. [7], there are two parts in a shadow: self-shadow and cast shadow. The self-shadow is the part of an object which is not illuminated by direct light. The cast shadow is the region projected by the object in the direction of direct light. Our objective is to detect the cast shadow from the object. Although the formation of the cast shadow depends on environment factors, we have observed that there are several generic features that can be used to guide the proposed shadow detection method. The observations are as follows:

Observation 1: The research in Ref. [7] states that the ratio between pixels when illuminated and the same pixels under shadows can be roughly linear. Experiments found that the ratio ranges from 1 to 2.5;

Observation 2: A shadow always makes the region it covers darker, but does not change its hue significantly;

Observation 3: A shadow cast on a background often lowers the saturation of the points.

3.2 Shadow detection and suppression

The HSV color space corresponds closely to the human perception of color [7] and reveals more accuracy in distinguishing shadows than the RGB color space. The presented algorithm for shadow detection involves a two-stage process. First, the distortion of brightness for each pixel between the incoming frame and the reference image is computed to abstract the candidate pixels of moving region. The process is performed according to the following equation:

$$F(x,y)=\begin{cases}1, & \left|I^{\mathrm{V}}(x,y)-B^{\mathrm{V}}(x,y)\right|\geqslant 2*\sigma(x,y),\\ 0, & \text{otherwise},\end{cases} \tag{2}$$

where $\sigma(x,y)$ is the mean value of the distortion of brightness for the pixel at the position (x, y), and it is computed as follows:

$$\sigma(x,y,i) = \max(\sigma_{\min}, \alpha|I(x,y,i) - B(x,y)| + (1-\alpha)\sigma(x,y,i-1)), \quad (3)$$

where a minimum distortion value $\sigma_{\min}(x,y)$ is introduced as the noise threshold to prevent the background measurements from remaining strictly constant over a long period, which will in turn result in the distortion value decreasing below a minimum.

Second, the hue and saturation information of the candidate pixels, as well as the division of brightness between the pixels and the corresponding pixels in background image, are combined to determine the shadow points. Consequently, the shadow points are classified by the following decision procedure:

$$S(x,y) = \begin{cases} 1, & \alpha \leqslant \dfrac{B^{V}(x,y)}{I^{V}(x,y)} \leqslant \beta \wedge D^{H}(x,y) \\ & \leqslant \tau_{H} \wedge I^{S}(x,y) - B^{S}(x,y) \leqslant \tau_{S}, \\ 0, & \text{otherwise}, \end{cases} \quad (4)$$

where α and β are respectively the upper bound and lower bound of brightness division; τ_H and τ_S are selected threshold values used to measure the distortion of the hue and saturation between the background image and the current observed image. The distortion of the hue is computed as follows:

$$D^{H}(x,y) = \min\left(\left|I^{H}(x,y) - B^{H}(x,y)\right|, 360 - \left|I^{H}(x,y) - B^{H}(x,y)\right|\right). \quad (5)$$

Note that the division image is multiplied by a prefix k in our implementation to increase the scale sensitivity of the results, which makes further threshold operation more reliable and easier.

4 Experimental results

Both indoor and outdoor scenes of the proposed algorithm on video sequences are made available at http://cvrr.ucsd.edu/aton/shadow/index.html. In the experiment, the corresponding parameters are configured as follows:

$$\sigma_{\min} = 10,\ \alpha = 1,\ \beta = 3,\ k = 50,\ \tau_H = 20,\ \tau_S = 0. \quad (6)$$

The results of the running algorithm applied to the three sequences, Intelligent room (Fig. 3), Laboratory (Fig. 4), and Campus (Fig. 5) are shown respectively. In Fig. 3, the shadow pixels are correctly detected even if the shadow is too slight to be seen by human eyes. For the other indoor testing sequence, Laboratory, the algorithm fails to completely abstract the foreground pixels from the background due to significant similarities of the color information between the background scene and the moving individuals. Some background pixels are misclassified to foreground points suffering from high noises, as can be seen in Fig. 4. The result of the outdoor sequence, Campus, is shown in Fig. 5. It turns out to be a satisfying result for people detection and shadow suppression, although there are high noises and disturbance.

To systematically evaluate a shadow detector, it is useful to identify the following two important quality measures: good detection (low probability of misclassifying a shadow point) and good discrimination (the probability of classifying non-shadow points as shadow should be low, i.e., low false alarm rate). The first one corresponds

Fig. 3 Intelligent room

Fig. 4 Laboratory

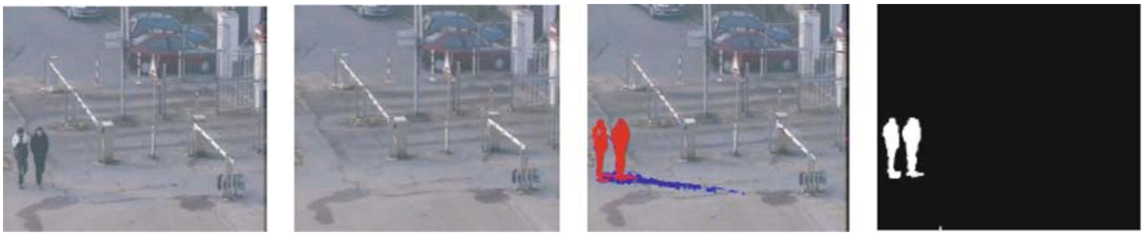

Fig. 5 Campus

to minimizing the false negatives (FN), i.e., the shadow points classified as background/foreground, while for good discrimination the false positives (FP) i.e., the foreground/background points detected as shadows, are minimized. In Ref. [3], two metrics for moving object detection evaluation are proposed: the detection rate (DR) and the false alarm rate (FAR). Assuming TP as the number of true positives (i.e., the shadow points correctly identified), these two metrics are defined as follows:

$$DR = \frac{TP}{TP + FN}, \; FAR = \frac{FP}{TP + FP}. \tag{7}$$

We apply these two metrics to make a quantitative evaluation of the presented shadow detection algorithm. To compute the evaluation metrics described above, the ground truth for each frame is necessary. It was obtained by segmenting images with an accurate manual classification of points in foreground, background, and shadow regions. The quantitative evaluation results are reported in Table 1.

Table 1 Experimental results of quantitative evaluation

sequence	sequence type	noise level	DR/%	FAR/%
Intelligent room	indoor	low	91.62	11.59
Laboratory	indoor	medium	89.38	12.46
Campus	outdoor	high	90.54	10.84

5 Conclusions

A novel background initiation and maintenance method for efficient indoor and outdoor people detection and a robust shadow detector by exploiting HSV color information is proposed. An experiment with indoor and outdoor sequences is employed to test the performance of the algorithm. A quantitative metric is further used to establish a more objective evaluation. The experimental results prove that the algorithm is efficient and robust.

References

1. Kalman K P, von Brandt A. Moving object recognition using an adaptive background memory. In: Proceedings of Time-varying Image Processing and Moving Object Recognition II. Netherlands: Amsterdam, 1990, 289–296
2. Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Colorado: Fort Collins. 1999, 2: 246–252
3. Prati A, Mikic I, Trivedi M M, et al. Detecting moving shadows: algorithms and evaluation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(7): 918–923
4. Chung Y C, Wang J M, Chen S W. A vision-based traffic light detection system at intersections. Journal of Taiwan Normal University: Mathematics, Science and Technology, 2002, 47(1): 67–86
5. Massey M, Bender W. Salient stills: Process and Practice. IBM Systems Journal, 1996, 35(3&4): 557–573
6. Toyama K, Krumn J, Brumit B, et al. Wallflower: principles and practice of background Maintenance. In: Proceedings of the 7th IEEE International Conference on Computer Vision, 1999, 1: 255–261
7. Herodotou N, Plataniotis K N, Venetsanopoulos A N. A color segmentation scheme for object-based video coding. In: Proceedings of IEEE Symposium on Advances in Digital Filtering and Signal Processing, 1998, 25–29