An Illumination Invariant Change Detection Algorithm†

Jianguang Lou, Hao Yang, Weiming Hu and Tieniu Tan
National Laboratory of Pattern Recognition (NLPR),
Institute of AutomationChinese Academy of Sciences,
Beijing, 100080 P. R. China
{jglou, hyang, wmhu, tnt}@nlpr.ia.ac.cn

Abstract

In this paper, a homomorphic filtering based change detection algorithm is proposed to detect moving objects from light-varing monocular image sequences. In our approach, a background model is first constructed, and background subtraction is applied to classify image pixels into background or foreground. We utilize illumination invariant local components to model the background, which are obtained using homomorphic filtering. Threshold for every pixel in the image can be selected automatically to accommodate the change of lighting. In addition, the connectivity information is integrated into the background-foreground classification process by Bayesian estimation. Experimental results show that the presented approach works well in the presence of heavy moving shadows and illumination variance.

1. Introduction

Robust and efficient motion detection is an important preprocessing step to solve many problems in the area of computer vision, such as shape from motion, visual surveillance and object recognition, etc. One of the most common approaches to this problem is background subtraction which provides the complete feature data in the current image. The background image describes the stationary portion of the scene, and moving objects can be identified as those regions of pixels in the image that differ significantly from the background. Therefore, a robust background model is a fundamental component in this kind of approaches. Unfortunately, background modeling is hard and time-consuming, and is not well solved yet.

In the past twenty years, color or intensity based approach [1, 2, 4, 5, 7, 8, 9] and range based approach [10, 11] are proposed for change detection. In [3], the authors integrate the range and color cues into change detection. Because range based algorithm is limited to multi-view image sequences as the range information is usually obtained by stereo vision, color or grayscale based approaches are more suitable for monocular vision applications. The algorithm presented in this paper is precisely of this kind.

To reduce the sensitivity to variations in lighting, many approaches based on statistical models have recently been proposed. A standard color based method for establishing an adaptive background model is temporal averaging, and the background model is estimated by regions which are similar to the current scene except where motion occurs. However, it is not robust to such scenes with multiple moving objects particularly when they move at very low speed. Some pixel-wise methods were proposed which separate the image sequence into independent pixel processes. In [1, 5, 7], each pixel is modeled by a single Gaussian process. Ridder et al. [2] modeled each pixel of background with a Kalman filter, which made their system to fit the lighting changes. Sun et al. [1] used histogram to restore background from an image sequence. Furthermore, Stauffer et al. [4] have discussed a mixture Gaussian model for each pixel of background and the background model is updated over time. The difference between the current image and the reference image is evaluated by thresholding. In some applications, because of the presence of moving shadows, this thresholding operation would give wrong result by misclassifying moving shadows as parts of moving objects. This will make the subsequent processing (for example, object recognition) more erroneous. In [8], the author proposed an approach to handle this problem, but the threshold is still selected

† This work is funded by Chinese NSFC.

empirically. How to select threshold automatically remains to be an open problem.

In this paper, we establish a robust change detection framework that is flexible to deal with variations in lighting and the presence of moving shadows. We first use illumination invariant characteristics to describe the model of background. After this, the connectivity information was integrated into the background-foreground classification operation using Bayesian estimation. Our approach can automatically select threshold for every pixel in the image.

2. The method

The basic steps of background subtraction algorithm are listed as follows:

Background modeling: It uses a statistic model to represent the background image. In many existing algorithms, RGB values are directly used to model the background. It is not robust to the change of the lighting. In our approach, we combine color cues and brightness information to construct the model. The color components in our approach are preprocessed by homomorphic filtering to avoid the influence of lighting changes. Generally, the scene illuminance varies smoothly over space and locates at low frequency part in frequency domain, whereas In addition, reflection components locate at relatively high frequency part. According to this, we can separate them by homomorphic filtering. Once we obtain the reflection components which are invariant to lighting changes, we can use them to model the background for eliminating the influence of lighting changes.

Pixel classification: It classifies the pixels into foreground or background regions. Thresholding operation is usually applied in this step. Thus, the selection of an appropriate threshold value for every image pixel is of great importance. In our approach, the connectivity information is integrated into this step, because most moving objects always manifest themselves as compact connected regions.

2.1. Background modeling

The image of an object is generated by an incoming illumination which is reflected by the surfaces of the object. Considering "white" illumination, the color components of the image pixel can be calculated by [6]:

$$I(j)=e(j)\cdot\left[m_b(\vec{n},\vec{s})\cdot k_c(j)+m_s(\vec{n},\vec{s},\vec{v})\cdot f(j)\right] \quad (1)$$

where e is the spectral power distribution(SPD) of the incident light, $m_b(\vec{n},\vec{s})$ and $m_s(\vec{n},\vec{s},\vec{v})$ denote the geometric dependencies on the body and surface reflection, k_c is a compact expression depending on the sensor, *f* denotes the specular reflection parameter, and *j* is the pixel-index. We simplify Eqn. (1) into:

$$I(j)=l(j)\cdot r(j) \quad (2)$$

In Eqn. (2), $r(j)$ is an illumination invariant component and describes the reflection characteristic of the object surface. It is only determined by the material of the object surface. $l(j)$ is an illumination-dependent component. We try to separate $r(j)$ from $I(j)$. In order to do this, Equation (2) was first transformed into Eqn.(3) by using logarithm:

$$\ln I(j)=\ln l(j)+\ln r(j) \quad (3)$$

In most practical surveillance applications, it is justifiable to assume that the scene illuminance varies smoothly over space. Since the new image model is required to be invariant not only to the global change of lighting but also to smooth variations of the light distribution, the illuminance component should be acquired as local as possible. Therefore, a local 3x3 Gaussian low-pass filter is employed in our framework as described below. This process is known as homomorphic filtering. As mentioned above, it can separate reflection components from the image.

$$\ln l = G_{3\times3}\otimes \ln I$$

and $r' = exp(\ln I-\ln l)$ (4)

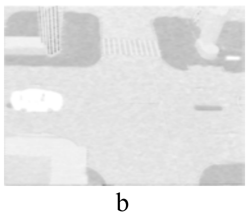
where G3x3 is an Gaussian function, and I denotes the image.

Color information is a simple but very important pixel-based feature for change detection. In many practical applications, the foreground has different color from the background. It is reasonable to integrate chromatic cues in order to obtain more robust performance. The rgb reflection characteristics of the background are calculated respectively.

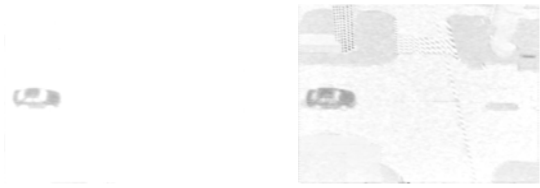
$$r_r' = exp(\ln I_r-\ln l)$$
$$r_g' = exp(\ln I_g-\ln l)$$
$$r_b' = exp(\ln I_b-\ln l) \quad (5)$$

Figure 1 shows an example for these reflection components. Significant difference can be seen between the background and foreground in the reflection images.

a



b



c d

Figure 1 the color reflection component. (a) the original image, (b) the red reflection component, (c) the green reflection component, (d) the blue reflection component.

Sometimes the color reflection components obtained by Eqn. (4) are different from the reflection components we have defined in Eqn. (2) by a scale factor, so they are denoted as r'. To reduce the influence of the difference between r and r', we should normalize them to form an unit vector [r_r r_g r_b]. In the r_r r_g r_b 3D space, all of these vectors locate on a spherical surface centered at the origin.

Assume that the pixel process belonging to the background is a set of Gaussian processes. Every background image pixel is modeled by a 4-tuple $< E_R, \sigma_R, l, \sigma_1 >$, where E denotes the expected illuminance invariant color vector [Err Erg Erb], $\sigma_R = [\sigma_r \quad \sigma_g \quad \sigma_b]$, l is the brightness component and σ_1 denotes the standard deviation of l.

To estimate these parameters, the method which has been proposed in our previous work[1] is employed in this paper. For any single pixel in an image sequence, we can find that it belongs to the background at most of the time, and is only occasionally occluded by moving objects. The history of each pixel can be represented by a set of histograms. The expected values of the pixel Gaussian processes can be estimated by the peak values of the histograms, and the standard deviations can also be obtained. It is preferable that there are illumination variances in the training images for obtaining good model parameters.

2.2. Pixel Classification

In the following, M denotes the change mask, which consists of a binary label m(j) for each pixel j on the current image grid. Each label m(j) either takes the value $m(j) = F$ ('foreground'), if pixel j is considered as part of moving a object, or the value $m(j) = B$ ('background'), if pixel j is hypothesized as background. As a special case of a Bayesian classification, we try to estimate the change mask M that can maximize its posteriori probability $P(M \mid I, I_b)$ by given the current image I and the reference image I_b. For each pixel j, the decision rule is then given by

$$\frac{P(m(j) = F \mid I(j))}{P(m(j) = B \mid I(j))} \overset{F}{\underset{B}{\gtrless}} t \tag{6}$$

with a global decision threshold t, which can be determined by setting the false alarm rate of the classifier. In this paper, we select t=3, and this decision rule will select the one that has the highest posteriori probability. Using Bayes' theorem, the decision rule can be rewritten as:

$$\frac{P(I(j) \mid m(j) = F)}{P(I(j) \mid m(j) = B)} \overset{F}{\underset{B}{\gtrless}} \frac{P(m(j) = B)}{P(m(j) = F)} \tag{7}$$

where $P(m(j) = B)$ and $P(m(j) = F)$ are the a-priori probabilities for pixel j's classification. In reality, we have no a prior knowledge with the respect to the expected change masks. When the first image of the image sequence is evaluated, no information is available on the value of $P(m(j) = B)$ and $P(m(j) = F)$. Thus, two steps are carried out to classify the pixels.

First step:

We use the following decision rule: A pixel in the current image is 'foreground', if the pixel has significant reflection characteristics different from the expected values in the reference image; it is selected as 'background', if the pixel has similar reflection components to the same pixel in the background image.

$$\begin{cases} m(j) = B, & \text{if} |r(j) - E(j)| < 3 \mid \sigma \mid \\ m(j) = F, & \text{otherwise} \end{cases} \tag{8}$$

where the selection of the factor 3 is in response to the characteristics of Gaussian distribution. If the data are normally distributed, 99.7% of the data are within 3 σ of the mean.

Second step:

After this step, pixels in the current image have been roughly divided into two sets of pixels. One is a collection of foreground pixels, denoted as C1. In this step, we refine the detection result, while the connectivity information is integrated into pixel classification. Because we already have obtained an initial collection C1 in the first step, the distributions of the parameters for foreground pixels can be estimated. In other words, we can calculate $P(I(j) \mid m(j) = F)$ in Eqn.(7). Furthermore, moving objects always manifest themselves as compact regions with smooth shape, and false alarms appear typically as small and scattered regions. We want to fill the small holes in image regions of the moving objects, and also to reject false alarms.

Connectivity is one of the cues which can be adopted in this problem. Aach etc. [12] described the change mask as samples from two-dimensional Gibbs /Markov random fields. For simplicity, our approach is a heuristic method based on the perception that the possibility of a pixel belonging to 'foreground' will increase if the number of its neighbors belonging to 'foreground' increases. Figure 2 shows an 8-neighbors neighborhood. In fact, the size of the neighborhood of a pixel should be adjusted according to the size of the foreground in the image. In some applications, 5x5 or 7x7 neighborhood may be used. For example, we adopt 5x5 neighborhood in the example described in Figure 5. For simplicity, only an 8-neighbors neighborhood is discussed here.

In Figure 2, horizontally or vertically adjacent pixels and diagonally adjacent pixels have different effects on pixel j because of the difference between their distances from pixel j. It is assumed that $A = D/\sqrt{2}$.

A	D	A
D	j	D
A	D	A

Figure 2. 8-Neighbors of pixel j

Furthermore, the value of $\frac{P(m(j)=B)}{P(m(j)=F)}$ can be estimated by:

$$\begin{cases} \frac{P(m(j)=B)}{P(m(j)=F)} = 1/100, \; if \; B_D = B_A = 0 \\ \frac{P(m(j)=B)}{P(m(j)=F)} = 100, \; if \, F_D = F_A = 0 \\ \frac{P(m(j)=B)}{P(m(j)=F)} = \frac{B_D\sqrt{2}+B_A}{F_D\sqrt{2}+F_A}, otherwise \end{cases} \quad (9)$$

where B_D and B_A are the number of pixels which have been classified as background in j's neighbors D and A respectively. F_D and F_A are the number of pixels which have been classified as foreground in j's neighbors D and A respectively. When $F_D = F_A = 0$ or $B_D = B_A = 0$, $\frac{P(m(j)=B)}{P(m(j)=F)}$ is estimated by the global a prior probability of a pixel being a foreground pixel in the image. In this paper, we set it as Eqn. (9), because in many applications, the foreground pixels occupy less than 1/20 of the image.

We integrate the ratio of probability for every pixel into decision rule (7), and reclassify all pixels at least one of whose neighbors has a class label different from its label after the first step. By using this method, the threshold can be automatically selected for every pixel in the current image. The second step can also iterate for many times in order to obtain a better result. Figure 3 (g) shows the effect of this step, and some false alarms are removed.

In reality, when the current image is not the first image in video sequence, the conditional probability $P(I(j)\,|\,m(j)=F)$ can be estimated from the previous frames. We need not to implement the first step by setting the change mask of previous frame to the initial change mask of the current frame, because it is reasonable to assume that the change mask of the current frame is similar to the previous one. An acceptable foreground mask can be obtained without iterating the second step.

Now, we further classify the background pixels into three classes using brightness information.

- ***Original background*** if the brightness component of the pixel is also similar to the corresponding pixel in the reference image.
- ***Shadow*** if the brightness of the pixel is lower than the same pixel in the reference image.
- ***Highlighted background*** if the brightness of the pixel is higher than the same pixel in the reference image.

As stated in [6], segmentation only with illuminance invariant color vector will "provide poor results in areas where intensity is small (i.e. within 5% of the total intensity range)". It is because the noise would constitute a large percentage of the pixel value, when the RGB values of the pixel were low. In reality, there might be a moving object that contains some very low RGB pixels, and these pixels will always be misclassified as Moving Shadow. On the other hand, some object pixels in background regions of low RGB values will probably be misclassified as Highlighted Background. In this case, brightness thresholding operation is employed for classification.

2.3. Background Updating

According to Equ. (1), the reflection components can be influenced by the variance of the spectrum of illumination source and the direction variance of the light source. For example, the spectrum of the sunlight in early morning is different from that in noon because of the dispersion by the atmosphere. This will lead to little difference on the reflection component. It will be more robust if we can update the background model over time. Our method is similar to [4]. The parameters of the background model are updated as follows

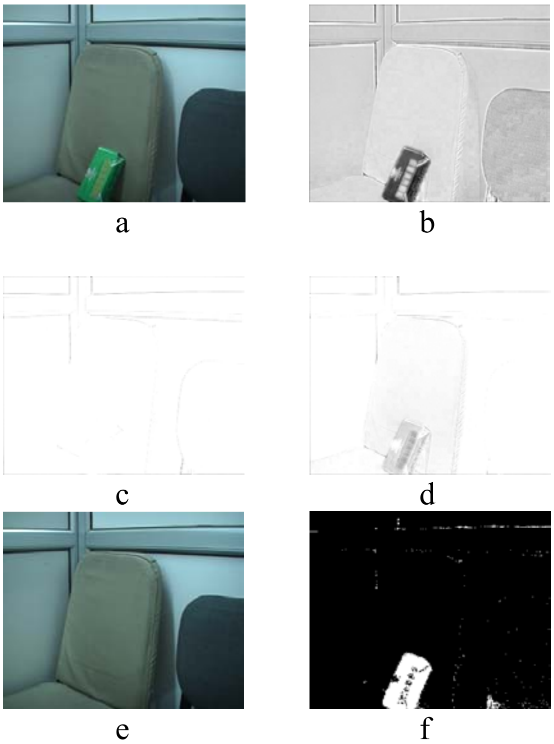
$$E_t = (1-\rho)E_{t-1} + \rho(R_t - E_{t-1})$$

$$\sigma_{r,t} = (1-\rho)\sigma_{r,t-1} + \rho(R_t - E_t)^T (R_t - E_t)$$

where updating parameter ρ is determined by the speed of lighting change.

3. Experimental results

In order to evaluate the performance of our approach, some experiments have been carried out. The image sequences showed are sets of true color images with 360x288 resolution which are captured by a PanasonicTM NV-DX100 digital video camera fixed on a tripod. In our experiments, our approach is performed on three video sequences. Because that the illumination component is eliminated by homomorphic filtering as described above, the background model we have obtained is not sensitive to the change of illumination. It can work well under lighting change and in the presence of shadows. The preliminary experimental results are shown in Figure 3, Figure 4 and Figure 5 respectively, from which we can see the effectiveness of our approach.



g h

Figure 3. An example of object detection, (a) a frame in an indoor video sequence, and the lighting has changed largely, (b) the image of r_r, (c) the image of r_g, (d) the image of r_b, (e) the reference image, (f) the detection result of the first step, (g) the result after integrating connectivity constraint, (h) the detection result using conventional background subtraction

By comparing Fig.3 (g) and Fig.3 (h), we can see that our approach can correctly classify heavy shadows and objects, and clearly outperforms the conventional subtraction method.

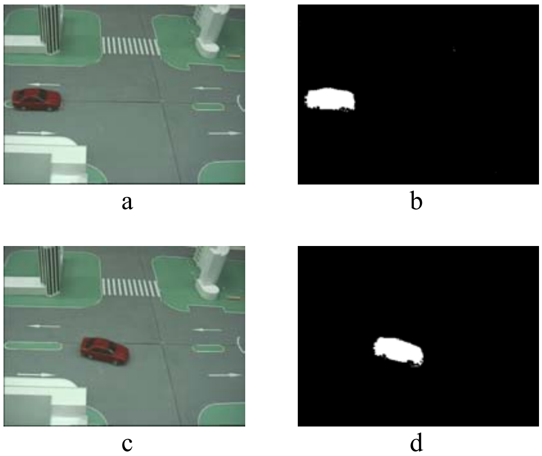


Figure 4. A traffic surveillance sequence. (a) The 11th frame of the sequence, (b) our detection result of (a), (c) the 18th frame of the sequence, (d) our detection result of (c).

In Figure 4, a traffic surveillance sequence is demonstrated. We use this change detection method in a traffic surveillance system. When the lighting condition changed from (a) to (c), our approach can detect the moving car robustly.

Another example of indoor scene is shown in Figure 5, from which we can see that the detection result is highly satisfactory, especially in handling moving shadows.

4. Conclusion

In this paper, we have presented a novel change detection approach based on background subtraction. The background is modeled using illumination invariant local components which are obtained using homomorphic filtering. Our approach can automatically select thresholds for every pixel in current image by integrating the connectivity constraint into classification. Experimental results show that this new method performs well even under lighting changes and in the presence of shadows.

5. References

[1] H. Z. Sun, T. Feng and T. N. Tan, "Robust extraction of moving objects from video sequences", *Proc. of the Fourth Asian Conference on Computer Vision*, Vol.2, Jan. 2000, pp. 961-963.

[2] C. Ridder, O. Munkelt, and H. Kirchner, "Adaptive Background Estimation and Foreground Detection using Kalman-Filtering", *Int. Conf. on Recent Advances in Mechatronics*, UNESCO Chair on Mechatronics, 1995, pp. 193-199.

[3] G. Gordon, T. Darrell, M. Harville, J. Woodfill, "Background estimation and removal based on range and color", *Int. Conf. of Computer Vision and Pattern Recognition*, 1999, (http://www.ai.mit.edu/~trevor/papers/1998-071/).

[4] C. Stauffer, W. E . L. Grimson, "Adaptive background mixture models for real-time tracking", *Int. Conf. of Computer Vision and Pattern Recognition,* Vol.2, 1999, pp. 246-252.

[5] S. Huwer, H. Niemann, "Adaptive Change Detection for Real-time Surveillance Applications", *Workshop on Visual Surveillance in ECCV2000*, 2000.

[6] T. Gevers, A. W. M. Smeulders and H. Stokman, " Photometric Invariant Region Detection", *Proc. of British Machine Vision Conference (BMVC)* , 1998, pp. 659-669.

[7] C. Wren, A. Azarbayejani, T. Darrell and A. Pentland, "Pfinder: Real-Time Tracking of the Human Body", *IEEE Trans. PAMI*, vol 19, no 7, 1997, pp. 780-785.

[8] J. Stauder, "Segmentation of Moving Objects in Presence of Moving Shadows", *Proc. of VLBV 97*, Linkoeping, Sweden, July 1997.

[9] A. Cavallaro and T. Ebrahimi, "Change detection based on color edges", *Proc. of IEEE Int. Symposium on Circuits and Systems (ISCAS-2001)*, Sydney (Australia), 6-9 May, 2001.

[10] Y. Ivanov, A. Bobick and J.Liu, "Fast Lighting Independent Background Subtraction", *Proc. of the IEEE Workshop on Visual Surveillance*, Bombay, India, January 1998, pp. 49--55.

[11] C. Eveland, K. Konolige and R. C. Bolles, "Background Modeling for Segmentation of Video-Rate Stereo Sequences", *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, June 1998, pp. 266--271.

[12] T. Aach, A. Kaup and R. Mester, "Change Detection in Image Sequences Using Gibbs Random Fields: A Bayesian Approach", *IEEE Proc. Int. Workshop on Intelligent Signal Processing and Communications Systems*, Sendai, Japan, Oct. 1993, pp. 376-381.

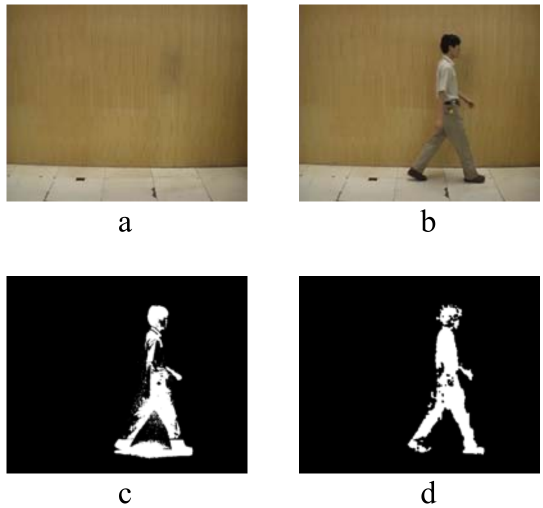


Figure 5. A change detection result for human tracking. (a) The reference image, (b) one frame of a human tracking video, (c) the detection result by directly using RGB values, (d) the result of our approach