MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11 - 13, 2002, Nara- ken New Public Hall, Nara, Japan

3—10 Adaptive Background Estimation for Object Tracking

Ryunosuke Itoh*, Yoshio Iwai*, Masahiko Yachida*
*Graduate School of Engineering Science
Osaka University

Abstract

Tracking people has received considerable attention by computer vision researchers. Interest is motivated by the broad range of potential applications such as personal identification, human-machine interaction, and automated surveillance. We previously proposed a method for extracting moving region by subtraction from background images. Background images were estimated by using UD factorized Kalman Filter. In order to use a Kalman Filter, good initial parameters are needed for estimation and were manually given in the previous work. In this paper we propose an improved method for estimation of background image and for automatic parameter initialization. We also show experimental results using real images to test the performance of our proposed method.

1 Introduction

It is important for a detection method to be robust for illumination changes and shadows in order to detect moving objects stably in outdoor scenes. Many methods have been proposed for removing illumination effects. Ohta uses a linear illumination model and chi-square test for detecting outliers[1]. Chromatic difference is frequently used because shadows do not effect chromatic component[2, 3]. Other statistical approaches are also used for determination of outliers[4, 5]. Tao et. al use a neural network for removing shadows[6]. Comparative evaluation of each method have also been proposed [7].

In order to address the problems of illumination changes and casting shadows simultaneously, we use a linear model for expressing a pixel value. We decompose a pixel value into two components: a surrounding light and a direct light. Shadows can be expressed by a blending coefficient of these two components. Illumination changes caused by clouds is considered as intensity changes of illumination because it is acceptable that clouds uniformly reflect all frequency.

In this paper, we propose a method for detection of objects which cast shadows in an outdoor scene by modeling a pixel value as a linear combination of a surrounding light and a direct light. The method can also deal with slow color change of sunlight by using a Kalman filter because we use it for estimation of the two background components. We conduct experiments on real images in an outdoor scene to evaluate our proposed method

*Address: 1-3 Machikaneyama-cho, Toyonaka, Osaka 560-8531, Japan. E-mail: {itoh, iwai, yachida}@yachi-lab.sys.es.osaka-u.ac.jp

2 Adaptive Background Model

In this paper, we deal with images in outdoor scenes, so we assume that a light source in a scene is only sunlight. Let intensity of the surrounding light be $\mathbf{L}_{sj}$, intensity of the direct sunlight be $\mathbf{L}_{dj}$, and the blending coefficient matrix of an object surface be $\mathbf{r}_j$. We assume that a pixel value $\mathbf{E}_j$ is given by the following equation:

$$\mathbf{E}_j = \mathbf{L}_{sj} + \mathbf{r}_j \mathbf{L}_{dj}, \qquad (1)$$

where $\mathbf{r}_j$ is a diagonal matrix including albedo and specular reflection of an object surface. By using the above equation, we can estimate a parameter $\mathbf{r}_j$ when $\mathbf{L}_{sj}$ and $\mathbf{L}_{dj}$ are known in advance. $\mathbf{L}_{sj}$ and $\mathbf{L}_{dj}$ are estimated from an observation by using a Kalman filter. $\mathbf{r}_j$ is estimated from an observation by assuming that a neighbor region of the observation has uniformly the same value of reflection.

3 Estimation Algorithm

We describe the proposed algorithm for detection of moving objects and estimation of the background parameters: $\mathbf{r}_j$, $\mathbf{L}_{sj}$, and $\mathbf{L}_{dj}$.

3.1 Outline of Process

The outline of our method is shown in figure 1. At first, input images are divided into some blocks and each block is classified into the background block and the non-background block by performing subtraction and erosion. The illustration of the block classification process are shown in figure 2. White and black blocks in the figure express background blocks and non-background blocks, respectively.

All $\mathbf{r}_j$ of pixels in background blocks are calculated from the components, $\mathbf{L}_{sj}$ and $\mathbf{L}_{dj}$, of the background image and an input image $\mathbf{L}_j$ by using equation 1. However, $\mathbf{L}_{sj}$ and $\mathbf{L}_{dj}$ cannot be determined at the beginning of process, so the initial values are determined by a method described in the next section.

After calculation of $\mathbf{r}_j$, $\mathbf{L}_{sj}$ and $\mathbf{L}_{dj}$ are updated by using a Kalman Filter. Then shadow removal and object detection are performed.

3.2 Estimation of Initial Background Parameters

When we use a Kalman filter for background estimation, we need to take care of the initial parameters of the

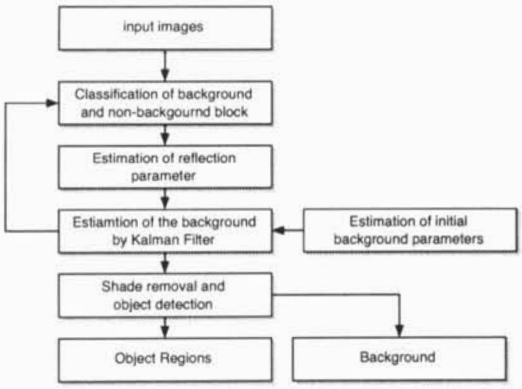
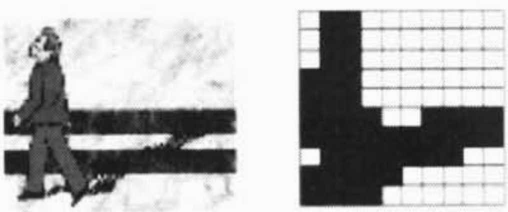


Figure 1: System flow of extraction process



(a) input image (b) classification result

Figure 2: Classification of input image

model because incorrect determination in the early stage of process causes worse final results.

First, the system calculates the median value of each pixel $\mathbf{L}_{med}$, the maximum and minimum intensity values of each pixel, T_{max} and T_{min}, for the last N frames.

From equation 1, $\mathbf{r}_j$ is nearly equal to 0 in dark regions in an input image. $\mathbf{L}_{med}$ is regarded as the ratio of RGB components. So we can assume the following equation:

$$T_{min} \cdot \mathbf{L}_{med} \simeq \mathbf{L}_s \cdot T_{med}, \tag{2}$$

where T_{med} is the intensity of the median value $\mathbf{L}_{med}$. Therefore, we can estimate the initial value of $\mathbf{L}_s$ by using the following equation:

$$\mathbf{L}_s = \frac{T_{min}}{T_{med}} \cdot \mathbf{L}_{med}. \tag{3}$$

$\mathbf{r}_j$ can be estimated from the intensity T_{Lj} at $\mathbf{r}_j = 0$ and the intensity T_{Hj} at $\mathbf{r}_j = 1$ by using the following equation:

$$\mathbf{r}_j = \frac{T_j - T_{Lj}}{T_{Hj} - T_{Lj}}, \tag{4}$$

where T_j is intensity of pixel j in an input image. In fact, T_{Lj} and T_{Hj} cannot be automatically estimated, so we use T_{max} and T_{min} in place of them and we get:

$$\mathbf{r}_j = \frac{T - T_{min}}{T_{max} - T_{min}}, \tag{5}$$

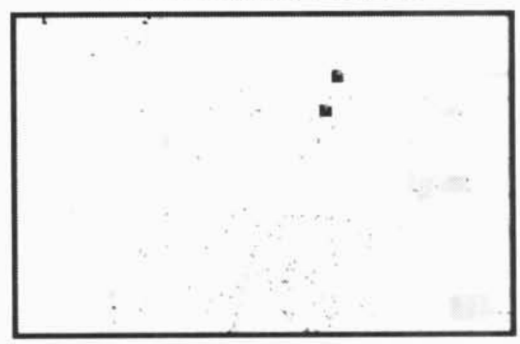
or we can get the following equation:

$$\mathbf{r}_j = \frac{T}{T_{max}}, \tag{6}$$

if we assume that T_{Lj} is equal to 0. In experiments, we use equation 6 because of good accuracy as shown in figure 3. $\mathbf{L}_{dj}$ is estimated by the following equation:

$$\mathbf{L}_d = (\mathbf{L} - \mathbf{L}_s) / \frac{T}{T_{max}}. \tag{7}$$

(a) initialization by equation 5



(b) initialization by equation 6

Figure 3: Subtraction results using different initial values

When an observed object is illuminated by a surrounding light, in other words, when clouds block the sunlight, input images are hardly considered as be pitch-black if the sensitivity of a camera is correctly adjusted. When the sunlight is not blocked by clouds, input images are not saturated if a camera has enough dynamic range not to be saturated. However, real cameras have the noise caused by dark current and they are saturated by specular reflection. We, therefore, define the upper and lower limits of $\mathbf{L}_{sj}$ and $\mathbf{L}_{dj}$ in order to reduce effects of dark current noise and saturation noise.

3.3 Updating Parameters by Kalman Filter

The background components, $\mathbf{L}_{sj}$ and $\mathbf{L}_{dj}$, of each pixel in the background block are estimated with the values, $\mathbf{r}_j$, described in the previous section by using a Kalman Filter. The system equation is as follows:

$$\mathbf{X}_t \equiv (\mathbf{L}_{sj}^T, \mathbf{L}_{dj}^T)^T, \tag{8}$$

$$\mathbf{X}_{t+1} = \mathbf{X}_t + \mathbf{N}(0, \sigma_1^2), \tag{9}$$

$$\mathbf{L}_t = \begin{pmatrix} 1 & [\mathbf{r}_j]_1 & & & & \mathbf{0} \\ & & 1 & [\mathbf{r}_j]_2 & & \\ \mathbf{0} & & & & 1 & [\mathbf{r}_j]_3 \end{pmatrix} \mathbf{X}_t + \mathbf{N}(0, \sigma_2^2),$$

where N is the Gaussian noise, σ_1 and σ_2 are standard deviations of the Gaussian, and $[\mathbf{r}_j]_i$ is the i-th diagonal component of the matrix. The feedback gain (Kalman gain) is calculated at each pixel by using the equation derived from the above system equation.

3.4 Shadow Removal and Object Detection

Shadow removal and object detection are performed by calculating chroma difference between object color and the background model, $\mathbf{L}_{sj}$ and $\mathbf{L}_{dj}$. The chroma difference vector $\mathbf{D}$ as shown in figure 4 is calculated by using the following equation:

$$\mathbf{D} = (\mathbf{L} - \mathbf{L}_s) - (\mathbf{L} - \mathbf{L}_s, \mathbf{L}_d - \mathbf{L}_s)(\mathbf{L}_d - \mathbf{L}_s), \quad (10)$$

where $(,)$ is the inner product operator. The chroma difference of a shadow region is small because the sunlight is blocked by an object and $\mathbf{r}_j$ is nearly equal to 0. Therefore, we classify pixels into object region and background region by using the following rules:

$$\begin{cases} \text{background} & \|\mathbf{D}\| \leq T, \\ \text{object} & \text{otherwise.} \end{cases} \quad (11)$$

The value of the threshold T is empirically determined.

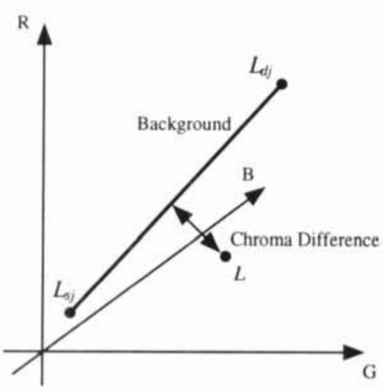
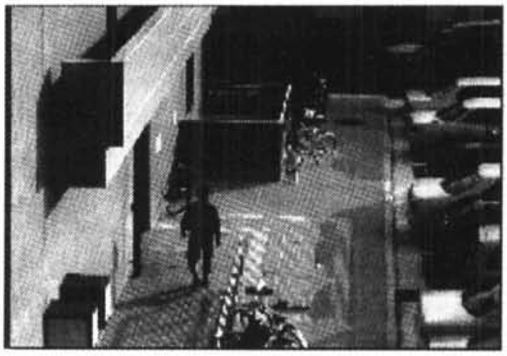


Figure 4: Chroma difference between object color and background model

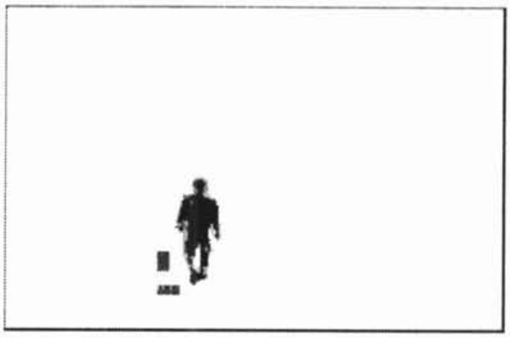
(a) 352nd frame



(b) 610th frame

Figure 5: Examples of input images

(a) 352nd frame



(b) 610th frame

Figure 6: Experimental results

4 Experimental Results

We conduct experiments using an SGI workstation (Onyx2, MIPS R12K, 400MHz). Figure 5 shows input images (720×486 pixels) captured at about 4 p.m. on October 24th, 2000 (fair occasionally cloudy).

4.1 Object detection and shadow removal

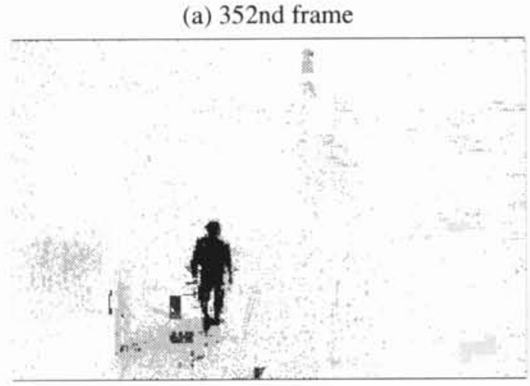
Figure 6 shows experimental results. The displayed areas in the figure are detected as a moving object.

From the result in figure 6 (b), background estimation failed after an object passed. The reason of the failure is that the determination of background and non-background regions is performed per block, but the background parameters are updated per pixel. Nevertheless few pixels in the block belongs to an object, the block is classified into a background region. So the background parameters of the pixels belonging to an object region are diverged because of a large Kalman gain in proportion to the chroma difference. Re-initialization is required in such a diverged region.

4.2 Comparison experiment

We also conduct an experiment to compare results of our method with those of median images. The median images are calculated from the last 20 frames. Figure 7 shows experimental results by using median images for subtraction. Gray regions express object regions detected by subtraction with median images and black regions express objects detected by our method. As clearly shown in figure 7, the shadow of an object on the building are not removed by using median images because the response of median is slow and cannot follow the changes caused by a moving object.

(a) 352nd frame



(b) 610th frame

Figure 7: Experimental results: median image

The upper and lower limits described in section 3.2 are empirical determined in this experiment, however these values are dependent on scenes. The proposed method can not treat with periodic changes like a leaf motion blowed by wind. In future work, we will improve the background model which can express such periodic changes.

5 Conclusion

We proposed a background subtraction method for detecting objects. The method can remove object's shadows and can deal with illumination changes occurred by clouds and other occluding objects. The method can also deal with the slow color change of sunlight (like a sunset) by using the Kalman filter. Our method is based on extrapolation of a surrounding light and direct ray of sunlight from observed pixel values, so the Kalman filter sometimes became unstable. In this paper, we use an UD factorized Kalman filter for stable estimation. In future work, we intend to improve our method which can deal with periodical color changes and can perform filtering more stably.

Acknowledgment

This work is partially supported by Japan Society for the Promotion of Science under Grant-in-Aid JSPS-RFTF 99P01404.

References

[1] N. Ohta. A statistical approach to background subtraction for surveillance systems. In *Proc. of ICCV*, pp. 481–486, Vancouver, Canada, July 2001.

[2] A. Elgammal, D. Harwood, and L. S. Davis. Non-parametric model for background subtraction. In *Proc. of IEEE ICCV'99 Frame-rate Workshop*, Greece, September 1999. IEEE.

[3] T. Horprasert, D. Harwood, and L. S. Davis. A statistical approach for real-time robust background subtraction and shadow detection. In *Proc. of IEEE ICCV'99 Frame-rate Workshop*, Greece, September 1999. IEEE.

[4] I. Mikic, P. Cosman, G. Kogut, and M. Trivedi. Moving shadow and object detection in traffic scenes. In *Proc. of ICPR*, volume 1, pp. 321–324, 2000.

[5] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *Proc. CVPR*, pp. 246–252, 1999.

[6] X. Tao, M. Guo, and B. Zhang. A neural network approach to the elimination of road shadow for outdoor mobile robot. In *Proc. Intl. Conf. on Intellight Processing System*, volume 2, pp. 1302–1306. 1997. IEEE.

[7] A. Prati, I. Mikic, R. Cuchiara, and M. Trivedi. Analysis and detection of shadows in video streams: A comparative evaluation. In *Proc. of CVPR*, volume 2, pp. 571–576, Hawaii, December 2001. IEEE.