Background Subtraction Technique based on Chromaticity and Intensity Patterns

Ariel Amato, Mikhail Mozerov, Iván Huerta, Jordi González, and Juan J. Villanueva.

Centre de Visio per Computador, Universitat Autònoma de Barcelona, 08193 Bellaterra Spain.
Centre Institut de Robòtica i Informàtica Industrial (UPC-CSIC) Edifici U, Parc Tecnològic de Barcelona. Llorens i Artigas 4-6, Barcelona 08028, Spain.

Abstract

This paper presents an efficient Real-Time method for detecting moving objects in unconstrained environments, using a background subtraction technique. A new background model that combines spatial and temporal information based on similarity measure in angles and intensity between two color vectors is introduced. The comparison is done in RGB color space. A new feature based on chromaticity and intensity pattern is extracted in order to improve the accuracy in the ambiguity region where there is a strong similarity between background and foreground and to cope with cast shadows. The effectiveness of the proposed method is demonstrated in the experimental results and comparison with others approaches is also shown.

1. Introduction

Detection of moving regions in a video scene plays an important role in computer vision field. Many application such as human detection and tracking, surveillance system, traffic monitoring are based on motion detection. Background subtraction is a simple and effective approach to detect and segment motion in video sequences. This technique has been actively investigated and applied for many researchers during the last years [1,2,3,4,5]. To obtain a precise segmentation, is necessary to overcome several difficulties like illuminations changes, moving shadows, camouflages etc. Consequently, a robust and accurate algorithm to segment moving object in different scenarios (indoor, outdoor) has to be developed. In this paper, we present several new modifications of similarity measurements that are used in background subtraction methods. These modifications concern with the chromaticity similarity and the neighborhood similarity. Instead of measuring the chromaticity similarity in the non-linear HSV color space we propose a simple measurement based on an angle between two color vectors. This modification helps us to solve the shadow suppression problem. The information provided by a single pixel, sometimes is not enough to distinguish between background and foreground due to the strong similarity that can exist between regions. In this case, we propose to measure similarity not between two color vectors, but between two sets of color vectors that form a neighborhood patterns. Such a modification helps us to improve the ability of our background subtraction methods to overcome the camouflage problem.

This paper is organized as follows. Section 2 introduces a brief of more classical Real-Time Background Subtraction approaches. Section 3 presents our method. In Section 4 experimental results are discussed. Concluding remarks are available in Section 5.

2. Related work

Haritaoglu et al. in W4 [4] use a model of background subtraction built from order statistics of background values during a training period. The background scene is modelled by representing each pixel by three values: its minimum and maximum intensity values and the maximum intensity difference between consecutive frames observed during this training period. Pixels are classified as foreground if the difference between the current value and the minimum and maximum values are greater than the
values of the maximal interframe difference. However, this approach is rather sensitive to shadows and lighting changes, since the only illumination intensity cue is used.

Horprasert et al. [5] implement a statistical color background algorithm, which use color chrominance and brightness distortion. The background model is built using four values: the mean, the standard deviation, the variation of the brightness and chrominance distortion. However, the chromaticity noise model in this paper is not correct. Indeed, according to the logic of the thresholds definition the chromaticity noise does not depend on the change of illumination, but such an assumption is wrong in general.

Kyungnam Kim et. al [3] use a similar approach, but they obtain more robust motion segmentation in the presents of the illumination and scene changes using background model with codebooks. The codebooks idea gives the possibility to learn more about the model in the training period The authors propose to cope with the unstable information of the dark pixels, but still they have some problems in the low and the high intensity regions.

Stauffer and Grimson [6] address the low and the high intensity regions problem by using a mixture of Gaussians to build a background color model for every pixel. Pixels from the current frame are checked against the background model by comparing them with every Gaussian in the model until a matching Gaussian is found. If so, the mean and variance of the matched Gaussian is updated, otherwise a new Gaussian with the mean equal to the current pixel color and some initial variance is introduced into the mixture.

Cucchiara et al. [1] use a model in Hue-Saturation-Value (HSV) and stress their approach in shadow suppression. The idea is that shadows change the hue component slightly and decrease the saturation component significantly. In the HSV color space a more realistic noise model can be done. However, this approach also has drawbacks. The similarity measured in the non-linear HSV color space, as a result the noise value seems to depend on the chromaticity value of the background color and such an assumption is not generally correct. Also, the color space transformation from RGB to HSV increases complexity of the approach and decreases the calculation error.

3. Proposed Algorithm

3.1. Similarity Measurements

To compare a background image with a current frame we use four similarity measurements. Those are:

- **Angular similarity measurement** $\Delta \theta$ between two color vectors in the RGB color space $p_1$ and $p_2$, which is defined as follows

$$
\Delta \theta(p_1, p_2) = \cos^{-1}\left(\frac{p_1 \cdot p_2}{|p_1||p_2|}\right).
$$

(1)

- **Intensity similarity measurement** $\Delta I$ between two color vectors in the RGB color space $p_1$ and $p_2$

$$
\Delta I(p_1, p_2) = |p_1| - |p_2|.
$$

(2)

With each of the described similarity measurements we associate a threshold function

$$
T \theta(\Delta \theta) = \begin{cases} 1 & \text{if } \Delta \theta > T^\theta, \\ 0 & \text{else} \end{cases}
$$

and

$$
T I(\Delta I) = \begin{cases} 1 & \text{if } |\Delta I| > T^I, \\ 0 & \text{else} \end{cases}
$$

(3)

where $T^\theta$ and $T^I$ are intrinsic parameters of the threshold functions of the similarity measurements.

To describe a neighborhood similarity measurement let us first characterize the index vector $x=(n,m)^T \in \Omega = \{0,1,\ldots,N; 0,1,\ldots,M\}$, which define the position of a pixel in the image. Also we need to name the neighborhood radius vector $w=(i,j)^T \in W = \{-W,\ldots,0,\ldots,W; -W,\ldots,0,\ldots,W\}$, which define the positions of pixels that belong to the neighborhood relative any current pixel. Indeed, the domain $W$ is just a square window around a chosen pixel.

- **Angular neighborhood similarity measurement** $\eta_\theta$ between two sets of color vectors in the RGB color space $p_{1,w}$ and $p_{2,w}$ ($w\in W$) can be written as

$$
\eta_\theta(p_{1,w}, p_{2,w}) = \sum_{w \in W} T \theta(\Delta \theta(p_{1,w}, p_{2,w})).
$$

(4)

where the functions $T \theta$ and $\Delta \theta$ are defined in Eq. (3) and Eq. (1) respectively.

- **Intensity neighborhood similarity measurement** $\mu_\theta$ between two sets of color vectors in the RGB color space $p_{1,w}$ and $p_{2,w}$ ($w\in W$) can be written as
\[
\mu_\theta(p_{\text{in}}^{l}, p_{\text{in}}^{z}) = \sum_{w \in \Omega} T_I(\Delta I(p_{\text{in}}^{l}, p_{\text{in}}^{z})), \tag{5}
\]
where the functions \(T_I\) and \(\Delta I\) are defined in Eq. (3) and Eq. (2) respectively.

With each of the neighborhood similarity measurements we associate a threshold function

\[
\begin{align*}
T\eta \theta(\eta \theta) &= \left\{\begin{array}{ll}
1 & \text{if } \Delta \theta > T^{\eta \theta} \\
0 & \text{else}
\end{array}\right., \\
T\mu I(\mu I) &= \left\{\begin{array}{ll}
1 & \text{if } \Delta I > T^{\mu I} \\
0 & \text{else}
\end{array}\right.,
\tag{6}
\end{align*}
\]
where \(T^{\eta \theta}\) and \(T^{\mu I}\) are intrinsic parameters of the threshold functions of the neighborhood similarity measurements.

### 3.2. Background Modeling

Our background model (BG) will be represented with two classes of components one we call running components (RC) and another we call training components (TC). The RC is a color vector in RGB space and only this component can be updated in running process. The TC is a set of fixed thresholds values obtained during the training.

The background model is represented by

\[
BG = \{\{p^h\}, \{T^\theta, T^I, T^{\eta \theta}, T^{\mu I}, W\}\}, \tag{7}
\]
where \(T^\theta\) is maxima of the chromaticity variation (temporal-base); \(T^I\) is maxima of the intensity variation (temporal-base); \(T^{\eta \theta}\) is the chromaticity pattern threshold (spatial-base); \(T^{\mu I}\) is the intensity pattern threshold (spatial-base); \(W\) is the half size of the neighbourhood window.

To obtain the background parameters in the definition of Eq. (7) the training process has to be performed. This first step consists of estimating the value of the RC and TC during the training period.

To initialize our BG we put the RC = \{\(p^h\)\} as the initial frame and these values are supported during all training process. To estimate \(T^\theta\) and \(T^I\) during the training period, we compute the intensity and chromaticity difference between a background image pixel and the related pixel in the current frame belonging to the training process

\[
\begin{align*}
T^\theta &= \max_{j \in \{1, 2, \ldots\}} \{\Delta \theta(p^h, p^j)\}, \\
T^I &= \max_{j \in \{1, 2, \ldots\}} \{\Delta I(p^h, p^j)\},
\end{align*}
\tag{8}
\]
where \(F\) is the number of frames in the training period.

In this paper we consider the simplified version of our algorithm. In this case the spatial-base thresholds \(T^{\eta \theta}\) and \(T^{\mu I}\) we put as 1 for the neighbourhood radius equaling 1 (the 3x3 square windows), and for all pixels in the frame. However, locally adaptive version of our approach allows estimating this parameter in each pixels like the temporal-based thresholds.

After the initialization has been done the following equations show how to obtain the TC values. These values are estimated in the **Running and Foreground Classification Process**.

**Our classification rules are enunciated in two steps:**

**Step One:** This step is concentrate on the pixels that have strong chromaticity and intensity differences with the background, it means that the following rule expression have to be TRUE or 1

\[
T \theta(\Delta \theta(p^h, p^j)) \cap T \eta \theta(\eta \theta(p^h, p^j)) = 1; \tag{9}
\]
where \(\gamma\) is an experimental scale factor for the training set thresholds and the value of this factor greater than 1. In this case the tested pixel is directly classified like foreground. Otherwise the classification will be done in the next step.

**Step Two:** Due to the angular similarity measurement is inaccurate for low intensity color vectors, we define a confidence threshold \(T'^\gamma\), where if the vectors are greater than \(T'^\gamma\) the classification will be done according to the rule of Eq. (10) otherwise according to the rule in Eq. (11).

\[
\begin{align*}
T \theta(\Delta \theta(p^h, p^j)) \cup T \eta \theta(\eta \theta(p^h, p^j)) &= 1; \tag{10}
\end{align*}
\]

\[
\begin{align*}
T I(\Delta I(p^h, p^j)) \cap (T I(p^h, p^j) > 0) \\
\cup T \mu I(p^h, p^j) &= 1; \tag{11}
\end{align*}
\]

**Updating Models:** The update of the background follows the next rule: put the current pixel value to the BG if this pixel is not classified as foreground.

### 4. Experimental Results

We applied the proposed algorithm in several videos scenes, under different conditions (indoor/outdoor). In order to evaluate the performance of the proposed technique we generated a ground-truth segmentation masks by manual segmentation. Due to the numbers of frames tested, from different sequences was considerably high, we choose 20 frames per scene which these were sampled in such way to maintain the same inter-frame distance from the beginning to the
end of the sequence and four different sequence has been evaluated. Using the same sequences we implements algorithms from others authors [3,4,5,6] to obtain a general comparison of our method.

Two quantitative expressions were utilized to evaluate segmentation process, False Positive Error (FPE) and False Negative Error (FNE). Eq (12). The FPE means that the background pixel was set as a Foreground and FNE the foreground pixels that were set as a Background.

\[ \text{Error(\%)} = \frac{\text{numbers of misclassification pixels}}{\text{numbers of correct foreground pixels}} \times 100 \] 

(a)  
(b)  
(c)  
(d)  

Fig. 1 Segmentation errors. a) FPE, b) FNE

Fig. 2 shows the comparison between our techniques and some well known methods. It can be seen that our method performs better in terms of segmenting camouflage areas and suppressing strong shadows.

5. Conclusions

We proposed a Background Subtraction algorithm whose effectiveness was demonstrated in the comparison with other methods. The background model combines spatial and temporal information based on similarity measure in angle and intensity. Also, a feature based on chromaticity and intensity pattern is extracted in order to resolve the ambiguity that exists between similar regions and to cope with moving shadows.

Acknowledgements

This work is supported by EC grants IST-027110 for the HERMES project and IST-045547 for the VIDI-video project, and by the Spanish MEC under projects TIN2006-14606 and CONSOLIDER-INGENIO 2010 MIPRCV CSD2007-00018. Jordi González also acknowledges the support of a Juan de la Cierva Postdoctoral fellowship from the Spanish MEC.

References