Research Article

Detection of Moving Objects with Fuzzy Color Coherence Vector

Yu-Long Qiao,1 Kai-Long Yuan,1 Chun-Yan Song,2 and Xue-Zhi Xiang1

1 College of Information and Communication Engineering, Harbin Engineering University, Building 21, 145 Nantong Street, Harbin 150001, China
2 College of Mechanical and Electrical Engineering, Northeast Forestry University, 26 Hexing Street, Harbin 150040, China

Correspondence should be addressed to Yu-Long Qiao; qiaoyulong@hrbeu.edu.cn

Received 18 March 2014; Accepted 25 July 2014; Published 13 August 2014

Academic Editor: Wanquan Liu

Copyright © 2014 Yu-Long Qiao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background subtraction is a popular method for detecting foreground that is widely adopted as the fundamental processing for advanced applications such as tracking and surveillance. Color coherence vector (CCV) includes both the color distribution information (histogram) and the local spatial relationship information of colors. So it overcomes the weakness of the conventional color histogram for the representation of an object. In this paper, we introduce a fuzzy color coherence vector (FCCV) based background subtraction method. After applying the fuzzy $c$-means clustering to color coherence subvectors and color incoherence subvectors, we develop a region-based fuzzy statistical feature for each pixel based on the fuzzy membership matrices. The features are extracted from consecutive frames to build the background model and detect the moving objects. The experimental results demonstrate the effectiveness of the proposed approach.

1. Introduction

With the increasing significance and popularity of digital multimedia applications such as video surveillance, object tracking, action understanding, anomaly detection, and gait recognition, the detection of moving objects in video frames occupies a critical role in the subsequent tasks. A successful detection system with great accuracy promotes the overall performance of the further processing. Therefore, various methods have been proposed for detecting moving objects in the past years. However, robust detection of moving objects in presence of complex backgrounds is yet a challenging problem. The basic detection method is based on frame difference. For consecutive frames of the video, it compares the intensities of the pixels at the same location and extracts the objects with an appropriate threshold. But this approach is sensitive to various disturbs, such as illumination changes, swaying vegetation, rippling water, and changes in the background geometry. When the number of frames in the image sequence is large and there is little change between consecutive frames, the background modeling and subtraction method is another solution for the detection problem [1].

The heart of a background subtraction method is background modeling that uses the new video frame to calculate and update a background model. This resulting background model provides a description of the entire background scene. The moving objects can be detected by identifying the pixels that are not consistent with the background model. Many different approaches have been proposed for modeling the background. The simple but often effective parametric approach is to model each pixel of a video frame with the single Gaussian distribution. The model parameters (mean and covariance) can be recursively updated with a simple adaptive filter [2]. This model works well for the static background. In the natural environments, even though the camera is fixed, the background appears to be dynamic because it always includes repetitive motions like swaying vegetation, rippling water, flickering monitors, and so forth. To model the dynamic background, the use of a single Gaussian becomes inappropriate and the mixture of Gaussians model (MoG) is preferable [3, 4]. When changes in the background are fast, the Gaussian assumption based modeling method is not suited. Therefore, a nonparametric approach was proposed for background modeling [5]. This method makes use of a general nonparametric kernel density estimation technique for
building a statistical representation of the scene background. A major drawback of this approach is that it ignores the time-series nature of the problem. Moreover, KDE requires training data from a sequence of examples that have a relatively "clean" background [6].

The pixel-based method is partially inspired by an observation that the background model will be well established if the background is static and there are sufficient training frames. Therefore, for the dynamic texture background and real-time applications, these pixel-based methods do not perform well. To overcome their limitation, researchers pay attention to the region-based methods that usually divide an image into (overlapping or nonoverlapping) blocks and calculate block features, from which we can build the background model and detect the foreground. Mason and Duric [7] extracted a feature vector based on the edge histogram to describe a block and then to detect the moving objects. Heikilä and Pietikäinen [2] proposed a block-based background subtraction method based on the local binary pattern (LBP) texture measure. Each image block was modeled as a group of weighted adaptive LBP histograms. Zhang et al. [8] proposed the spatiotemporal local binary pattern (STLBP) to model dynamic textures. In another work [9], they furthermore used the covariance descriptor, defined based on various spatial and texture features, to efficiently suppress dynamic textures in the background. Recently, several authors have explored the adoption of fuzzy approaches to deal with different aspects of detecting moving objects. Sigari et al. [10] proposed a fuzzy version of running average method (FG) for background subtraction. Maddalena and Petrosino [11] introduced a fuzzy learning factor into the background model update procedure (FASOM). W. Kim and C. Kim [12] introduced a fuzzy color histogram (FCH) based block features for background subtraction. The background model is reliably constructed by computing the similarity between local FCH features with an online update procedure.

A color histogram provides only a distribution of the color. The FCH could attenuate color variations generated by background motions. However, color histogram is a very coarse image characterization so that images with similar histograms can have dramatically different appearances. Color coherence vector (CCV) does include both the color distribution information (histogram) and the local spatial relationship information. Therefore, this paper proposes a background subtraction method based on fuzzy color coherence vector that is derived from CCV by applying the fuzzy c-means clustering. The local features extracted from the fuzzy color coherence vector are used to build and update the background model and detect the moving objects.

This paper is organized as follows. In Section 2, we provide a brief overview of color coherence vector and introduce the fuzzy color coherence vector. Section 3 presents the background modeling and moving objects detection method. To demonstrate the effectiveness of the proposed background subtraction method, the experimental results on various image sequences will be shown and analyzed in Section 4. Section 5 draws a conclusion for the paper.

2. Fuzzy Color Coherence Vector

2.1. Color Coherence Vector. Color is an essential and important property of an image; thus, it has been explored in various image processing problems including background subtraction. A color histogram can be regarded as the probability density function that describes the probability for pixels in the image to belong to a color bin. It is computationally efficient and generally insensitive to small changes in camera position and geometric transformation such as rotation. However, a color histogram provides only a distribution of the color. It is a very coarse image characterization so that images with similar histograms can have dramatically different appearances.

In order to overcome the limitations of the conventional color histogram, Pass and Zabih [13] propose a histogram refinement method and give a new color characterization, color coherence vector (CCV), that includes both the color distribution information (histogram) and the local spatial relationship information. As shown in Figure 1, there are two images with the same color histograms. However, they may represent different objects. Because the descriptor CCV covers spatial relationship information of the color of pixels, the two images can be distinguished with their CCVs.

For a usual RGB image, there are three color channels, and each color level is 256, which is of heavy computational burden to obtain CCV. In addition, the human visual perception system does not match with RGB color space very well. Therefore, we first convert RGB into HSV color space. In general, taking into account computational complexity and perceptual representation performance, the HSV image will be quantized into a gray image with $m$ bins (levels). This paper adopts the quantization schemes: 4 hues, step 90 (315, 45, 135, 225, 315), 4 saturations, step 0.25 (0, 0.25, 0.5, 0.75, 1.0), and 8 values, step 0.125 (0, 0.125, 0.25, ..., 1.0), which will generate a gray image $I_{HSV}$ with 128 bins [14]. In order to improve the detection performance, we also quantize value (V) into 128 bins resulting in another gray image $I_{V}$ with 128 bins. The images $I_{HSV}$ and $I_{V}$ will be combined to detect the moving objects.

In the CCV framework, a pixel within a given color bin is classified as either coherent or incoherent. A coherent pixel is a part of a sizable contiguous region, while an incoherent pixel is not. Each CCV represents this classification for a color bin in the image. To determine the coherent pixel, we compute connected components using 8-connected neighbors with a given color bin. A pixel is coherent if the size of its connected
component exceeds a threshold; otherwise, the pixel is incoherent.

Suppose that \( \alpha_1 \) and \( \beta_1 \) are the numbers of coherent and incoherent pixels in the \( ith \) color bin of the histogram, respectively. Then the color histogram of the image with \( m \) color bins is described as \( (\alpha_1 + \beta_1), (\alpha_2 + \beta_2), \ldots, (\alpha_i + \beta_i), \ldots, (\alpha_m + \beta_m) \) and the CCV of the image is defined as \( (\alpha_1, \beta_1), (\alpha_2, \beta_2), \ldots, (\alpha_i, \beta_i), \ldots, (\alpha_m, \beta_m) \) \[13, 15\]. The judging threshold \( th \) for the coherence pixel is defined as follows:

\[
th = \gamma \left( N_{i1} + N_{i2} + \cdots + N_{ij} + \cdots + N_{im} \right),
\]

where \( N_{ij} \) denotes the number of pixels of the \( jth \) individual coherent regions for the \( ith \) color bin, \( n_i \) is the total number of the regions, and \( \gamma \) is a weighting coefficient to control the accuracy of detection and is obtained from practical training. As shown in Figure 1, there are six pixels with green color in each image resulting in the same histograms. However, if we set the threshold \( th \) to be 2, then the corresponding elements in the CCVs of the two images are \((3, 3)\) and \((6, 0)\), respectively.

When an object is moving into a region, not only the local color histograms but also the local spatial relationships of the color of this region are changing. The CCV includes both the distribution and spatial relationship information of the color. Hence, the CCVs of \( L_{HSV} \) and \( L_V \) should give better clues for moving objects detection.

### 2.2. Fuzzy Color Coherence Vector

Based on the observation that background motions (dynamic texture) do not make severe alterations of the scene structure, W. Kim and C. Kim \[12\] adopt the fuzzy color histogram (FCH) to attenuate color variations generated by background motions and improve moving objects detection performance. The CCV is the refinement of the color histogram by considering the local spatial relationship. Therefore, this paper combines fuzzy c-means clustering with CCV and introduces the fuzzy color coherence vector (FCCV) for background subtraction.

After obtaining the conventional CCVs of \( L_{HSV} \) and \( L_V \), \( (\alpha_1^{HSV}, \beta_1^{HSV}), \ldots, (\alpha_m^{HSV}, \beta_m^{HSV}) \) and \( (\alpha_1^{V}, \beta_1^{V}), \ldots, (\alpha_m^{V}, \beta_m^{V}) \), we apply the fuzzy c-mean algorithm to the combined coherence subvector \( \langle \alpha, \alpha_2, \ldots, \alpha_{mc,m} \rangle \) \( \langle \alpha = (\alpha_1^{HSV}, \alpha_1^{V}), \ldots, \alpha_m^{HSV}, \beta_1^{HSV}, \ldots, \beta_m^{HSV}, \beta_1^{V}, \ldots, \beta_m^{V} \rangle \), \( k = (i-1) \times m + j \) and obtain a membership matrix \( U^C \) and \( d \) clusters. The matrix \( U^C \) describes the membership degree of every element \( \alpha_i \) belonging to each cluster. For conducting the FCM algorithm, given the initial membership values \( u_{ij} \) that are randomly generated and subject to the condition \( \sum_{j=1}^{d} u_{ij} = 1 \), \( 1 \leq j \leq m \times m \), the initialization values of clustering centers \( \alpha_i \) are obtained by using (2), and then they are updated iteratively as \[12\],

\[
u_{ij} = \frac{1}{\sum_{k=1}^{d} \left( \frac{\| \alpha_j - \alpha_i \|}{\| \alpha_j - \alpha_k \|} \right)^{1/(r-1)}} \]

For \( 1 \leq i \leq d \), \( 1 \leq j \leq m \times m \),

\[
c_i = \frac{\sum_{j=1}^{m \times m} u_{ij} \alpha_j}{\sum_{j=1}^{m \times m} u_{ij}} \quad \text{For} \ 1 \leq i \leq d,
\]

where \( r \) is a constant and set to 2, which controls the spread degree among the fuzzy clusters. The clustering process stops when the maximum number of iterations 100 is reached, or when the objective function improvement between two consecutive iterations is less than the specified threshold. The same clustering method is applied to the combined incoherence subvector \( \langle \beta_1, \beta_2, \ldots, \beta_{mc,m} \rangle \) \( \langle \beta_k = (\beta_1^{HSV}, \beta_1^{V}), \ldots, (\beta_k^{HSV}, \beta_k^{V}) \rangle \) and obtains a membership matrix \( U^I \).

Once we obtain the two fuzzy membership matrices, each pixel of the HSV image can be endowed with the membership values to coherence clustering centers and incoherence clustering centers, respectively, by using the corresponding gray values of \( L_{HSV} \) and \( L_V \) images. Both the membership matrices \( U^C \) and \( U^I \) for the first video frame are calculated and stored in advance. They could be treated as two look-up tables. For each pixel of remaining video frames, we can obtain its membership values to coherence and incoherence clustering centers by using gray values of \( L_{HSV} \) and \( L_V \) images and table-look-up method without computing membership matrices.

In order to construct the background models, we define the region-based feature FCCV that can be easily built by referring to the stored membership matrices \( U^C \) and \( U^I \). For a video frame, we convert it into HSV color space and then quantize the HSV image into two grey images as the first frame. For a pixel at the position \( g \), it is assumed that the grey values of the \( L_{HSV} \) and \( L_V \) images are \( i \) and \( j \) \((1 \leq i, j \leq m)\), respectively. Then the membership value \( u^C_{gi} \) of this pixel belonging to the \( th \) coherence clustering centers is the \( kth \) element \( [U^C]_{thk} \) \((k = (i-1) \times m + j)\) of the matrix \( U^C \), while the membership value \( u^I_{gi} \) to the \( hth \) incoherence clustering centers is the \( hth \) element \( [U^I]_{thk} \) of the matrix \( U^I \). The region-based features vector (FCCV) \( F_g \) of the \( gth \) pixel is defined as follows:

\[
F_g = [F^C_g, F^I_g] = [f^C_{g,1}, f^C_{g,2}, \ldots, f^C_{g,d}, f^I_{g,1}, f^I_{g,2}, \ldots, f^I_{g,d}]
\]

\[
f^p_{g,h} = \sum_{z \in W_g} u^p_{g,z} \quad p = C, I, h = 1, \ldots, d,
\]

where \( W_g \) indicates the position set of the region centered at the location \( g \), \( u^p_{g,z} \) is the membership value of the pixel located at the position \( z \) to the \( hth \) coherence (incoherence) clustering center. By comparing the feature vectors of the same pixel location in the consecutive frames, we can detect the moving objects and update the background model.

### 3. Background Subtraction Methods

The background model BF is initialized with the feature vector FCCV of each pixel of the first frame of the image sequence. For the current frame, we compute the FCCV for every pixel and compare it with the FCCV of the corresponding position of the background model by using some similarity measure. Then the pixel is determined as
Mathematical Problems in Engineering

4 Mathematical Problems in Engineering

Input

Convert and quantize

If \( i = 1 \)

Calculate CCV and membership values

Calculate the FCCV for each pixel

Initialize the background model \( BF(1) \)

If \( i = 2, \ldots, T \)

Calculate the FCCV for each pixel

Detect the moving targets with \( BF(i - 1) \)

Update the background model

End

FIGURE 2: Flowchart of the proposed algorithm.

background \((O_g = 0)\) or foreground \((O_g = 1)\) objects as follows:

\[
O_g = \begin{cases} 
0, & D(F_g^i, BF_g^i) > \tau, \\
1, & \text{otherwise},
\end{cases}
\]

where \(D(\cdot, \cdot)\) is a similarity measure defined as

\[
D(F_g^i, BF_g^i) = \frac{\sum_{h=1}^{d} \min\left( f_{gh}^i, b_{gh}^i \right)}{\max\left( \sum_{h=1}^{d} f_{gh}^i, \sum_{h=1}^{d} b_{gh}^i \right)} \ast \varepsilon
\]

\[
+ \frac{\sum_{h=1}^{d} \min\left( f_{gh}^i, b_{gh}^i \right)}{\max\left( \sum_{h=1}^{d} f_{gh}^i, \sum_{h=1}^{d} b_{gh}^i \right)} \ast (1 - \varepsilon),
\]

where \(\varepsilon\) is an empirical parameter to integrate the incoherent and coherent features to compare the similarity of the two descriptions. That is to say, when the higher similarity exists between a pixel of the current frame and the background model, this pixel is implied to be a background. Otherwise, the lower similarity suggests that the pixel should belong to the moving objects.

In the dynamic background, some change will take place in the background as time goes on. In order to take these changes into account and build a sound background model, the model should be online updated with the detection results of the current frame. The scheme for the \(g\)th pixel is adopted as the following method [16]:

\[
BF_g^i(i) = \begin{cases} 
(1 - \lambda) BF_g^i(i - 1) + \lambda F_g^i(i), & \text{if } O_g = 0, \\
BF_g^i(i - 1), & \text{if } O_g = 1,
\end{cases}
\]

where \(\lambda\) is the updating rate belonging to the coverage \([0, 1]\). The updated background model \(BF(i)\) is used to detect the objects of the \((i + 1)th\) frame.

The background subtraction algorithm is shown in Figure 2 and summarized as follows.

(A) Initialization Phase (for the first video frame, \(i = 1\))

(a) Convert RGB into HSV color space and quantize it into two gray images \(L_{HSV}^i\) and \(L_V^i\) with \(m = 128\) color bins.

(b) Calculate the color coherence vectors of \(L_{HSV}^i\) and \(L_V^i\) and then apply FCM clustering method to the combined coherence subvector and incoherence subvector to obtain membership matrices \(U_C^i\) and \(U_I^i\), respectively.

(c) Derive the FCCV for each pixel using (3).

(d) Initialize the background model \(BF(1)\) \((i = 1)\) with the FCCV calculated in step (c).

(B) Detection Phase (for the \(i\)th frame, \(i = 2, \ldots, T\))

(a) Convert RGB into HSV color space and quantize it into two grey images as the initialization phase.

(b) Calculate all feature vectors (FCCVs) for the \(i\)th frame using (3).

(c) Detect the moving object by masking each pixel with “1” and “0” using (4) and the background model \(BF(i - 1)\).

(d) Update the background model using (6).

4. Results and Discussions

In this section, we conduct the experiments on several benchmark image sequences. In order to fully demonstrate the detection performance, the results of the proposed algorithm are compared to typical background subtraction methods including frame difference (FD) [17], MoG [3, 17], FCH [12], FG [10, 17], and FASOM [11, 17]. The numbers of the quantized color bins and clustering centers are chosen as \(m = 128\) and \(d = 32\), respectively. The updating weighting rate \(\lambda\) is set to be 0.01. The size of the local window for extracting FCCV is 5 \(\times\) 5 pixels. In order to be fair, no postprocessing is used to improve the background subtraction performance.

The benchmark image sequences include FT, WS, CT and CA [18], WT [19], CJ [20], CE, and FT2 [21]. The frame numbers of original image sequence and ground truths are listed in Table I. The selected original video frames, ground truths, and the results (foreground masks) of these methods are shown in Figures 3 and 4. It can be seen that the proposed
Figure 3: Experimental results (a) original frames, (b) ground truth, (c) FD, (d) MoG, (e) FCH, (f) FG, (g) FASOM, and (h) FCCV.
Figure 4: Experimental results (a) original frames, (b) ground truth, (c) FD, (d) MoG, (e) FCH, (f) FG, (g) FASOM, and (h) FCCV.
method provides a reliable background model so that the
dynamic background, such as waving leaves and turbulent
water, is almost separated from the moving objects. The
detection results suggest that the introduced method could
effectively find the moving objects and it is better than the
existing methods.

The foreground detection could be considered as binary
classification of each pixel. With the ground truth, we can use
the measures, recall and precision, to explain the correctness
of this classification. Thus, they could be used to evaluate the
performance of background subtraction methods:

\[
\text{recall} = \frac{\text{number of foreground pixels is correctly identified}}{\text{number of foreground pixels in the ground truth}},
\]
\[
\text{precision} = \frac{\text{number of foreground pixels is correctly identified}}{\text{number of foreground pixels is detected}}.
\]  

(7)

For the quantitative evaluation of our algorithm, we utilize a
composite measure, $F$-measure [22],

\[
F\text{-measure} = \frac{2 \cdot \text{recall} \cdot \text{precision}}{\text{recall} + \text{precision}},
\]  

(8)

to evaluate the effectiveness of our algorithm quantitatively. Higher values indicate better performance of the algorithm.

The $F$-measures of background subtraction algorithms for different image sequences are listed in Table 2. It is shown from the quantitative analysis that the proposed FCCV provides higher values on the average, which indicates better performance of the introduced method.

In our experiments, the foreground pixel is determined
by using (4) with the empirical parameter $\tau$. In order to
further evaluate the proposed algorithm, we also analyze the
detection performance with different values of the parameter $\tau$. The results for the image sequences WT, WS, and CA are shown in Figure 5. It can be seen that the detection performance has acceptable change when the parameter $\tau$ lies in the interval $[0.4, 0.5]$, from which we can choose a proper parameter value for detecting the moving objects.

5. Conclusions

The detection of moving objects in video frames occupies
a critical role in lots of applications. The traditional color
histogram provides important color distribution information
for moving object detection. However, there is a lack of the
local spatial relationship information of colors, which could
be captured by CCV. Therefore, this paper proposes a back-
ground subtraction algorithm based on FCCV using fuzzy $c$-
mean clustering and CCV, in which the introduced feature
vector FCCV for each pixel is used to build background
model and detect the moving objects. In addition, the fuzzy
property of the FCCV makes the algorithm robust to varying
changes of environment, so the method could detect the mov-
ing objects in the dynamic background. The experimental
results on some benchmark datasets demonstrate that the
approach achieves better detection performance than several
existing methods.

Conflict of Interests

The authors declare that there is no conflict of interests
regarding the publication of this paper.

Acknowledgments

This work is partially supported by the National Natural
Science Foundation of China 61371175, Postdoctoral Science-
Research Developmental Foundation of Heilongjiang
Province LBH-Q09128, and Fundamental Research Funds
for the Central Universities HEUCFQ1411. The authors also
gratefully acknowledge the helpful comments and sug-
gestions of the reviewers, which have improved the
presentation.
Table 2: F-measure.

<table>
<thead>
<tr>
<th></th>
<th>FT</th>
<th>WT</th>
<th>WS</th>
<th>CT</th>
<th>CA</th>
<th>CJ</th>
<th>CE</th>
<th>FT2</th>
</tr>
</thead>
<tbody>
<tr>
<td>FD</td>
<td>0.3726</td>
<td>0.7257</td>
<td>0.4797</td>
<td>0.6719</td>
<td>0.1133</td>
<td>0.5481</td>
<td>0.3186</td>
<td>0.1632</td>
</tr>
<tr>
<td>MoG</td>
<td>0.2160</td>
<td>0.5497</td>
<td>0.1582</td>
<td>0.3844</td>
<td>0.2712</td>
<td>0.5629</td>
<td>0.1952</td>
<td>0.3242</td>
</tr>
<tr>
<td>FCH</td>
<td>0.6341</td>
<td>0.9617</td>
<td>0.9050</td>
<td>0.7553</td>
<td>0.6466</td>
<td>0.4915</td>
<td>0.7989</td>
<td>0.6231</td>
</tr>
<tr>
<td>FG</td>
<td>0.5014</td>
<td>0.7492</td>
<td>0.6735</td>
<td>0.8460</td>
<td>0.1495</td>
<td>0.8239</td>
<td>0.3148</td>
<td>0.2753</td>
</tr>
<tr>
<td>FASOM</td>
<td>0.7849</td>
<td>0.9289</td>
<td>0.8702</td>
<td>0.8771</td>
<td>0.1712</td>
<td>0.8138</td>
<td>0.5679</td>
<td>0.5625</td>
</tr>
<tr>
<td>FCCV</td>
<td>0.7458</td>
<td>0.9839</td>
<td>0.8826</td>
<td>0.8198</td>
<td>0.7575</td>
<td>0.9131</td>
<td>0.8188</td>
<td>0.6866</td>
</tr>
</tbody>
</table>

References


