Moving Objects Detection Based on Bhattacharyya Distance Measurement

1Shahad Mahgoob Nafi and 2Sawseen Abdulhadi Mahmood
1College of Medicine, University of Baghdad, Baghdad, Iraq
2Department of Computer Science, College of Education, Al-Mustansiriyyah University, Baghdad, Iraq

Abstract: Motion detection in video sequences is considered as a critical issue in video surveillance systems. The key result of motion detection process is basically aims to extract and track the moving objects by segmenting the original frame into foreground part (moving objects) and background part in a video sequences. Practically, the moving objects may exhibit abnormal actions such as steady persons for while then moving. Meantime, the background may include periodic local actions such as waving arbor. Such intricacies may cause incorrect detection of moving objects based on the existing background subtraction approaches. In this study, we present a modified approach for moving objects detection based on Bhattacharyya distance measurement to detect the interaction between 2 sequence frames in a video sequences captured by static camera taking in consideration an indoor and outdoor scenes. A modified background subtraction method is adopted to extract the foreground parts in the video sequences (moving objects). The performance evaluation of the proposed approach was conducted using different video files taken from public video database with assist of several performance metrics. The experiments results proved the effectiveness of the proposed approach for moving objects detection.

Key words: Bhattacharyya distance, background subtraction, object detection, moving objects, detection, experiments

INTRODUCTION

Motion detection is one of the significant tasks in video scene understanding and processing of the system. It helps to take out the important information from scenes that are used in several computer vision applications, for example, object tracking, automatic video surveillance, classification and activity understanding. These make motion detection have a very high active research field in computer vision and its employment in automated visual and control surveillance system. Visual surveillance is the main technology to compact against crime, terrorism, public safety used in town centers, schools, transport networks and hospitals, proficient management of transport networks and public utility such as railway crossings and traffic light. In general, motion can be detected using different methods such as optic (detected video using cameras) infrared (employ passive and active sensors), sound and vibration (using magnetic measuring devices and magnetic sensors), radio frequency energy (microwave motion detection and radiography). But among these entire methodologies camera based method is usually intended for the improvement of computer vision based systems and for the detection of moving objects (Manchanda and Sharma, 2016). Moving object detection is the way to extract the moving part (region of interest) from all around (Hu et al., 2004). To take out the foreground from the background at the beginning camera captured the video and then decompose into frames, the pixels from these frames are categorized into 2 groups, one belongs to the foreground and the second belongs to the background (Parekh et al., 2014). There are several traditional methods which are used to detect motion.

These methods can be categorized as frame difference, optical flow temporal difference and background subtraction methods (Shaikh et al., 2014) these are the essential methods for motion detection, these methods can be used separately or in grouping of each other to get the best recognition of motion; first, frame difference presence the calculation of the difference between 2 or more sequential images (Bouthemy and Lalande, 1993) but there is a problem in this method that the detected objects are imperfect and unwell presented. Second, method is the optical flow (Beauchemin and Barron, 1995) which gives totally information about the movement but the challenging in the real-time implementation because these methods calculated
movement that is usually slow. Third, temporal difference
this method is built on the pixel wise difference of the
sequential frames. The result of this method does not
good when the detection of object is either moving too
fast or too slow and this is because of the little difference
between video frames and this lead to the explanation of
object is missing (Hu et al., 2004). Finally, background
subtraction the implementation concern in this method
is also unpretentious in which the background frame is
considered a reference, after that the difference between
background frame and current frame is calculated. The
resulting image acquired is the image of a movable object
to be represented as the foreground object (Athanassious
and Suresh, 2012).

Most motion detection algorithms used background
subtraction techniques, in general they have 3 stages.
First, it needs a static camera to develop a model of the
background. Second, it computes the difference between
current image and the background image; Finally, we must
apply the threshold operation to select if a pixel takes
place with the background or with the foreground.
Detecting a moving objects in a video frames have always
been a very essential research for several computer vision
researchers because there are still various challenges that
prevent the accurate and correct motion detection due to
several complicated matters like shape, illumination
change, speed, shadow and so forth. In the last years,
there are many motion detection methods have been
suggested.

Jain and Nagel (1979) proposed the most ordinarily
utilized method temporal differencing. This method takes
the current frame and compared it with the previous frame.
The compared image is then thresholded to separation out
foreground objects. This technique fails, if objects have
uniformly distributed intensity values and if object stay
not moving for more than a frame period (1/fps). There is
another similar method which takes current frame and
compared it with first frame. But this method fails if any
organizational or lighting changes happen.

Another methods built on Gaussian Mixture Models
(GMMs) (Stauffer and Grimson, 1999) are utilized as a part
of higher multifaceted nature scenes in which each pixel
is modeled as a mixture of Gaussian. This strategy is
competent to redundant movements from mess, long
period scene variations and illumination change. The
Self-Organizing Background Subtraction (SOBS) algorithm
(Maddalena and Petrosino, 2008) utilize neural system
structure for background model by learning motion design
in self-organizing technique. The model is strong for
moving background, slow light variety, cast shadow and
cover.

Kemoucheand Aouf (2009) suggested a new
technique to distinguish the moving objects by
incorporating the data in regards to the flow along with
the RGB color information to determine the background
form. Yatim et al. (2013) proposed a mechanized image
based frame work to recognize the motions of the human
in an intensive environment. At first using background
forming for extracting and naming the target object and
after that the centroid position of total objects are
collective to calculate the motion vector. Hu et al. (2015)
proposed a new technique for following and tracking the
moving item in a video frame using dynamic camera. At
first, the foreground object was separated with the
assistance of the image differences scheme and regions
related with the target object are acquired with the
assistance of incorporation scheme. At last with the
assistance of least bounding boxes they extracting
moving object which are then followed with Kalman filter
built on center of gravity. This method was good for real
time application and prepared for following the covering
objects.

MATERIALS AND METHODS

Various popular techniques initialize their model
using a sequence of frames such as the presented
research by Wang and Suter (2006). This technique seems
to be resistant to environmental noise this happens
because the approach provides enough information to
further sequential frames. But the model fails and results
in wrong detections if unexpected illumination changes
occur through initialization. In such a case it is required to
build another background model and the last whole model
is required to be rejected. Another failing in the model if
the length of analysis video is shorter than the training
video, this happens due to inadequate information. The
researchers in the presented research by Barnich and Van
Droogenbroeck (2011) suggested a solution for this
problem by initializing the background model from one
frame. Initializing the model done by choosing pixel value
in the spatial proximity. This leads to share the
information with neighborhood pixel as in the situation in
temporal distribution as one frame does not hold enough
information about color or the variation of illumination in
coming video sequences, thus, this process is liable to
noise problems.

In this research, we suggest a suitable solution for
this problem based on background subtraction, motion
detection and foreground subtraction techniques as
illustrated in the following subsections.
Moving object detection approach: The framework of the proposed moving object detection is shown in Fig. 1 and relies on three main steps:

- Step 1: background model construction
- Step 2: motion detection approach
- Step 3: foreground extraction approach

Background model construction: In this step, the first frame was taken as initial frame, then background frame needs to be updated in real time in order to update the background frame in order to avoid light changing and to extract the moving object accurately.

Motion detection: In this step, the human motion is detected based on the Bhattacharyya distance measure.

Foreground extraction approach: In this step, we use the background subtraction method for extracting the difference image and Otsu thresholding algorithm to calculate and determine the thresholding benchmarks to generate the binary image (binary mask process). Subsequently, the moving object can separated effectively with assist of foreground and background construction.

Background initialization: There are several ways to get the initial background image. Such that, put the first frame as the initial background or the background frame is constructed using the average of the first N frames or employ a background image sequences without the probability of dynamic object to evaluate the background model parameters. When we use a fixed camera, the simplest way is represented by adapting the first frame as background frame. In this way, this technique is effect significantly by outside location it is not good for changing in the environment with difficult background. The background frame configuration play an important role in the accuracy of object detection and updating the background image, a perfect background model is needed to extract the background accurately when there is a moving object in or out the background through the initialization. As the previous knowledge, we take first Frame (F1) of the input video sequence and supposed to be the initial background BC:

\[ BC(x, y) = F_1(x, y) \]  

(1)

Background construction model: The background model needs to be updated throughout video sequences in order to avoid background scene changing, moving objects appearance throughout the scene, noise, lighting variations and fixed object for a long time, etc. Therefore, this stage is so, crucial to find a suitable solution for these challenges in real time objects detection and tracking. In this study, the background frame model is constructed based on the Blind Background theory (EL Baf et al., 2008) which is denoted by:

\[ BC_{n+1}(x, y) = (1-\alpha) * BC_n(x, y) + \alpha C_n(x, y) \]

(2)

where, \( \alpha \) is a scalar value within the range (0, 1) determined from the experiments. In this research, \( \alpha = 0.8 \), \((x, y)\) is a grey pixel value in the current frame, \( BC_n(x, y) \) and \( BC_{n+1}(x, y) \) are the background value of current frame and next frame, respectively. In this research, the background model is constructed and updated at each 2 frames. The video sequences are acquired using fixed camera, so, the background model can stay relatively steady for a long period of time. Utilizing this method can effectively avoid the unexpected event of background like appearance of something sudden in the background that is not contained within the original background. Algorithm (1) illustrates the main steps of background model construction process.

Algorithm 1: (Background Construction):

Input: Video sample, frame-gap = 2, \( \alpha = 0.8 \)
Output: background
Motion detection: Human motion analysis and detection is the first task in computer vision applications. Human motion detection aims to identify the corresponding area of people from the overall image, which is a significant matter in human motion analysis systems followed by subsequent operations such as action recognition and tracking. There is a sequential process used in the motion detection algorithms and the foreground object extraction. In this study, the Bhattacharyya distance measure is utilized to measure the interaction between consecutive frames in real-time video which represents the key point of moving objects detection methods.

Bhattacharyya distance measure: Bhattacharyya distance (BH) one of the statistical measures widely used for evaluating. It first discovered by Bhattacharyya (1946). It's a variation-type measurement between 2 populations and applied in multiclass classification (Choi and Lee, 2003) and computer vision (Ahmern et al., 1998). The Bhattacharyya distance between 2 continuous probability distributions \( p_1, p_2 \) is defined as (AbdAllah and Shishmoni, 2013):

\[
\text{BH}(p_1, p_2) = -\ln \left( \text{BCH}(p_1, p_2) \right)
\]

where, \( \text{BCH} \) is the Bhattacharyya coefficient, a measure of the amount of overlap between 2 statistical samples. For the discrete probability distributions the Bhattacharyya coefficient will be:

\[
\text{BCH}(p_1, p_2) = \sum_{x} \sqrt{p_1(x) p_2(x)}
\]

And for continuous distribution it will be:

\[
\text{BCH}(p_1, p_2) = \int \sqrt{p_1(x) p_2(x)} dx
\]

Now, considering the special case of Gaussian distributions. Let, be 2 univariate Gaussian probability density functions where \( \mu_1, \mu_2 \) and \( \sigma_1^2, \sigma_2^2 \) and:

\[
p_1(x) = N(\mu_1, \sigma_1^2)
\]

\[
p_2(x) = N(\mu_2, \sigma_2^2)
\]

Then the Bhattacharyya coefficient is calculated as:

\[
\text{BCH}(p_1, p_2) = \sqrt{\frac{2\sigma_1 \sigma_2}{\sigma_1^2 + \sigma_2^2}} \exp \left( \frac{-\left( \mu_1^2 - \mu_2^2 \right)}{4(\sigma_1^2 + \sigma_2^2)} \right)
\]

Therefore, the BH is:

\[
\text{BH}_0(p_1(x), p_2(x)) = -\ln \text{BCH}(p_1(x), p_2(x))
\]

\[
\text{BH}_0(p_1(x), p_2(x)) = \frac{1}{2} \ln \left( \frac{\sigma_1 + \sigma_2}{2\sigma_1 \sigma_2} \right) + \frac{1}{4} \left( \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \right)
\]

Motion detection description: In this study, the description and implementation of the proposed motion detection algorithm is presented and illustrated which is composed of the following main steps:

Algorithm 2: The proposed motion detection algorithm:

Step 1: Read video file sequences

Step 2: Take the previous frame and current frame

Step 3: Convert the 2 sequence frames into gray level and calculate the histogram for both frames

Step 4: Calculate the Bhattacharyya distance between two successive frames as identified in Eq. 3

Step 5: If BH distance measure gives high difference between 2 successive frames and exceeds the given dynamic threshold then a motion is detected. Obviously, the estimation of an appropriate threshold level for the Bhattacharyya distance test is very critical to achieve the presented motion detection algorithm. Thus, the dynamic threshold level is adopted in this study based on the mean of the first 10 difference value obtained from BH-distance measure and weighted by a scalar value \( \beta \) determined from the experiments. In this research, \( \beta \) value is stand to (0.0001)

Additionally, to ensure accurate motion detection between the sequences frames, we adopt the Standard Deviation SD measure to discard the false positive status. The equation of the dynamic thresholding level is identified in Eq. 13:

\[
\text{Threshold} = \text{mean}(\text{differences buffer}) + \beta \times \text{SD}(\text{differences buffer})
\]
Algorithm 3: (Bhattacharyya distance measure):
Input: Video sample
Output: BH-distance
Begin
while not end of video file do
    current-frame - read frame from video sample
    if frame number greater than 1 then
        BH-distance = Bhat (previous-frame, current-frame)
    End if
    Previous-frame = current-frame
End while
End

Algorithm 4: (Motion detection algorithm):
Input: Video sample
Output: Motion (if moving = 1 then motion detection, if moving = 0 then no motion detection)
Begin
while not end of video file do
    BH-distance = Distance measurement ()
    thresholding dynamic-thresholding ( )
    if BH-distance <= thresholding then
        moving = 1
    else
        moving = 0
    End if
End while

Algorithm 5: (Dynamic thresholding):
Input: Video sample, β = 0.0001
Output: Thresholding
Begin
Buffer [ ] = 0 // Create one dim buffer [10]
while not end of video file do
    BH-distance = Distance measurement ( )
    if frame number is <10 then
        Right shifting for buffer
        buffer (last right element) = BH-distance
    else
        Right shifting for buffer
        buffer (last right element) = BH-distance
        mean-Gaussian distribution (buffer)
        SD (standard deviation) = Gaussian distribution (buffer)
        thresholding = mean (buffer) + beta + SD (buffer)
    End if
End while
End

Foreground extraction scheme: The foreground extraction procedure essentially based on 2 main parameters; the background subtraction and binary masking procedures. In order to extract the foreground (moving objects) from the video sequence, the absolute difference $D_i$ between the current frame and the background image is computed according to Eq. 14:

$$D_i(x, y) = |C_i(x, y) - B(x, y)|$$  \hspace{1cm} (14)

where, $i = 1, 2, ...$ is the frame index in the video sequence frames. The difference $D_i$ is calculated between the background image $B$ and the current frame $C_i$ when motion is detected as illustrated in algorithm 5. The binary masking procedure is performed according to the difference image $D$ and determining the optimum threshold value specified from Otsu method (Gonzalez and Woods, 2006).

Otsu method is mainly formulated to implement clustering-based image threshold, once 2 pixel classes’ background pixel and foreground pixel are supposed to be suitably differentiate as illustrated in Eq. 15. Otsu method computes the optimum threshold to split up the image into a foreground area, represented by $F$ and a background area represented by $G$. Through the minimization of within-class variation, threshold can decrease the classification errors. To compute the “Within-class variance” this is simply done by the sum of the 2 variances multiplied by their related weights:

$$\sigma^2_w(G) = w_c (g) \sigma^2_c (g) + w_f \sigma^2_f (g)$$  \hspace{1cm} (15)

$w_c (g)$ and $w_f (g)$ are the class probabilities at the intensity (g). And the 2 corresponding pixel classes $I_c$ and $I_f$ are the individual class differences. The formulas for element calculation are defined by AbdAllah and Shimshoni (2013). Threshold is computed in Eq. 16:

$$\Theta = \arg \min (\sigma^2_w (g))$$  \hspace{1cm} (16)

The thresholding process is performed according to Eq 17:

$$D(x, y) = \begin{cases} 1: D(x, y) \geq \Theta \\ 0: D(x, y) < \Theta \end{cases}$$  \hspace{1cm} (17)

Morphological operation: Morphological operations (Opening, closing) are a useful mathematical tool (Pratt, 1991). Mathematical morphology applied on the segmented binary image for smoothing the foreground region and describes the shape of area from image.
RESULTS AND DISCUSSION

The demonstration of the experimental results of the proposed motion detection algorithm is presented in this section based on public video database. The performance evaluation of the proposed motion detection algorithm is conducted using indoor and outdoor video samples taken from the adopted public data set (Change Detection.NET, 2014). The comparison operation is specified with the ground truth of the data set. Two types of video sequences were handled; the first type is a colored video while the second video type is captured by a thermal vision-based monitoring camera. From the experiments, we noticed the effective performance of the Bhattacharyya distance measure based motion detection algorithm which is showed an independence behavior regardless the video type (colored or not). In addition, the accuracy of edge classification is coupled with the closeness of tracked object to the camera. Thus, whenever the moving object is close to the camera, the Bhattacharyya distance based motion detection algorithm produces high differences.

We have adapted some performance measures to evaluate the performance of the proposed motion detection algorithm which discussed by Sokolova et al. (2006) and by Eisner et al. (2005). These measures are:

- **True Not Moving (TNM):** Based on the agreement of both the ground truth and the system result when there is no movement identified.

- **False Not Moving (FNM):** Based on the disagreement of both the ground truth and the system result when there is no movement identified.

- **True Moving (TM):** Based on the agreement of both the ground truth and the system result when there is movement identified.

- **False Moving (FM):** Based on the disagreement of both the ground truth and the system result when there is movement identified. For comparison purpose, 2 state of art (motion detection) algorithms are adopted, the first algorithm is Mixture of Gaussians (MOG1) and the second is adaptive background learning. The comparison results were obtained using testing program designed and implemented in MATLAB language program in order to exhibit the main differences between the results of the proposed method and the state of art methods.

<table>
<thead>
<tr>
<th>Video/Measurements tools</th>
<th>Proposed method</th>
<th>MOG1</th>
<th>Adaptive background</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Thermal-library</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TM</td>
<td>6408096</td>
<td>661123</td>
<td>4026459</td>
</tr>
<tr>
<td>FM</td>
<td>636601</td>
<td>20660</td>
<td>2547850</td>
</tr>
<tr>
<td>FNM</td>
<td>973595</td>
<td>6720568</td>
<td>3355232</td>
</tr>
<tr>
<td>TNM</td>
<td>41084228</td>
<td>41692769</td>
<td>39172979</td>
</tr>
<tr>
<td><strong>Baseline office</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TM</td>
<td>890911</td>
<td>4026459</td>
<td>793537</td>
</tr>
<tr>
<td>FM</td>
<td>285391</td>
<td>2547850</td>
<td>671305</td>
</tr>
<tr>
<td>FNM</td>
<td>205638</td>
<td>3355232</td>
<td>303012</td>
</tr>
<tr>
<td>TNM</td>
<td>18193300</td>
<td>39172979</td>
<td>17807386</td>
</tr>
<tr>
<td><strong>Baseline-pedestrians</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TM</td>
<td>96663</td>
<td>80285</td>
<td>99951</td>
</tr>
<tr>
<td>FM</td>
<td>11917</td>
<td>4595</td>
<td>168877</td>
</tr>
<tr>
<td>FNM</td>
<td>14486</td>
<td>30864</td>
<td>11198</td>
</tr>
<tr>
<td>TNM</td>
<td>9312321</td>
<td>9319643</td>
<td>9155361</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Video/Performance evaluation</th>
<th>Proposed method</th>
<th>MOG1</th>
<th>Adaptive background</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Thermal-library</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-measure</td>
<td>0.8884</td>
<td>0.1638</td>
<td>0.5770</td>
</tr>
<tr>
<td>Precision</td>
<td>0.9096</td>
<td>0.9593</td>
<td>0.6125</td>
</tr>
<tr>
<td>Recall</td>
<td>0.8681</td>
<td>0.89</td>
<td>0.5455</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.9847</td>
<td>0.9993</td>
<td>0.9389</td>
</tr>
<tr>
<td><strong>Baseline office</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-measure</td>
<td>0.7840</td>
<td>0.5770</td>
<td>0.6196</td>
</tr>
<tr>
<td>Precision</td>
<td>0.7574</td>
<td>0.6125</td>
<td>0.5417</td>
</tr>
<tr>
<td>Recall</td>
<td>0.8125</td>
<td>0.5455</td>
<td>0.7237</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.9846</td>
<td>0.9389</td>
<td>0.9637</td>
</tr>
<tr>
<td><strong>Baseline-pedestrians</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-measure</td>
<td>0.8798</td>
<td>0.8191</td>
<td>0.5261</td>
</tr>
<tr>
<td>Precision</td>
<td>0.8902</td>
<td>0.9459</td>
<td>0.3718</td>
</tr>
<tr>
<td>Recall</td>
<td>0.8697</td>
<td>0.7223</td>
<td>0.8093</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.9987</td>
<td>0.9995</td>
<td>0.9819</td>
</tr>
</tbody>
</table>

Obviously, in Fig. 5-7 the proposed method performance gives high level accuracy in comparison with MOG1 and adaptive background learning. Table 1, illustrates the output results of the measurement tools for the 3 methods. Based on the above measurements, we used the following performance indicators of the proposed motion detection algorithm:

\[
\text{Precision} = \frac{TM}{TM+FM}, \quad \text{Recall} = \frac{TM}{TM+FNM} \\
\text{Specificity} = \frac{TNM}{TNM+FM} \\
F\text{-score} = \frac{2 \cdot \text{Precision} \cdot \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}
\]

Table 2 gives a full description for the four performance evaluation metrics for video samples. An illustration of Bhattacharyya distance differences between the sequenced frames for a thermal-library video is presented in Fig. 2 with 1250 frames captured at thermal vision-based
movement at the beginning of video sequence, then the motion is detected at frame 861 based on BH-distance with dynamic threshold.

The proposed motion detection algorithm detects the object in real time when the person entering the room as we see BH-distance give high difference when the object approaching the camera. Then settle on a certain range when the human remains standing. Figure 4 shows the results for colored video sample (baseline pedestrians) with 800 frames. Also, there is no movement at the beginning of video sequence, then the motion is detected at frame 305 based on BH-distance (Fig. 5-7) demonstrates the original video frame with its foreground extraction in the ground truth and the proposed motion detection algorithm using 3 video data set, thermal-library video, baseline office video and baseline pedestrians video, respectively.
CONCLUSION

In our proposed method we defined a modified solution for human motion detection using BH-distance measure, background subtraction technique within indoor and outdoor environment to detect the moving object in a video sequence. The planned system was supported to eliminate processing requirements by eliminating analysis requests of unrelated background and reducing the processing time to detect motion in the successive frames. The results proved the efficiency of our method on scales of accuracy and low processing requirements. The proposed method can give perfect segmentation of the foreground-background in different situation and is study to illumination variation. We are supplementary working on other existing datasets to improve evaluation of the proposed algorithm.

REFERENCES


