Video Foreground Detection Based on Symmetric Alpha-Stable Mixture Models

Harish Bhaskar, Lyudmila Mihaylova, Senior Member, IEEE, and Alin Achim, Senior Member, IEEE

Abstract—Background subtraction (BS) is an efficient technique for detecting moving objects in video sequences. A simple BS process involves building a model of the background and extracting regions of the foreground (moving objects) with the assumptions that the camera remains stationary and there exist no movements in the background. These assumptions restrict the applicability of BS methods to real-time object detection in video. In this paper, we propose an extended cluster BS technique with a mixture of symmetric alpha stable (SαS) distributions. An on-line self-adaptive mechanism is presented that allows automated estimation of the model parameters using the log moment method. Results over real video sequences from indoor and outdoor environments, with data from static and moving video cameras are presented. The SαS mixture model is shown to improve the detection performance compared with a cluster BS method using a Gaussian mixture model and the method of Li et al. [11].

Index Terms—automatic object detection, background subtraction, segmentation, alpha stable distribution

I. INTRODUCTION

Moving object detection in video sequences represents a critical component of many modern video processing systems. The standard approach to object detection is Background Subtraction (BS), that attempts to build a representation of the background and detect moving objects by comparing each new frame with this representation [4]. A number of different BS techniques have been proposed in the literature and some of the popular methods include mixture of Gaussians [24], kernel density estimation [6], colour and gradient cues [9], high level region analysis [22], hidden Markov models [21], and Markov random fields [14]. Basic BS techniques detect foreground objects as the difference between two consecutive video frames, operate at pixel level and are applicable to static backgrounds [4]. Although the generic BS method is simple to understand and implement, the disadvantages of the frame difference BS is that it does not provide a mechanism for choosing the parameters, such as the detection threshold, and it is unable to cope with multi-modal distributions. One of the important techniques able to cope with multi-modal background distributions and to update the detection threshold makes use of Gaussian mixture models (GMMs). The model proposed in [24] describes each pixel as a mixture of Gaussians and an on-line update of this model. The larger Gaussian components correspond to the background and this is used to generate the background model. An algorithm for background modeling and BS based on Cauchy statistical distribution [13] is shown to be robust and adaptive to dynamic changes of the background scene and more cost effective than the GMM as it does not involve any exponential operation.

In [11] the foreground objects are detected in complex environments. The background appearance is characterised by principal features (spectral, spatial and temporal) and their statistics, at each pixel. However, the learning method in [11] requires ‘training’ since it relies on look up tables for the features and adapts them to the changes of environment. The CBS-SαS technique that we propose does not need such look up tables for the image features and is a cluster-based technique, which makes it different from [11]. According to our knowledge only one recent work [18] considers mixtures of SαS distributions for off-line data analysis and does not seem suitable for real-time object detection.

In this paper, we propose a novel cluster BS (CBS) technique based on SαS distributions, which technique we call CBS-SαS. The main contributions of the paper are threefold. Firstly, the BS process is performed at cluster level as opposed to pixel level methods that are commonly used [6], [24], [4]. The CBS-SαS method reduces significantly the clutter noise that arises due to slight variations in the pixel intensities within regions belonging to the same object. Secondly, due to their heavy tails, SαS distributions can help handling dynamic changes in a scene, and hence they model moving backgrounds and moving camera in a better way than the GMM. Results of modeling the background of a moving image sequence can be best obtained while operating with estimated values of the characteristic exponent parameter of the SαS distribution, rather than with fixed values corresponding to the Gaussian or Cauchy case. By estimating the parameters of the α stable distribution, the probability density function (PDF) of clusters of pixels can be faithfully represented and a reliable model of the background can be obtained. Thirdly, we show that a mixture of SαS distributions can represent well the multi-modality and guarantees reliable object detection. A wide range of tests is performed on indoor and outdoor environment, on data from a static and moving cameras.

The remaining part of the paper is organised as follows. Section II presents the proposed CBS-SαS technique. A comparison of the CBS-SαS with a CBS-GMM BS and

Harish Bhaskar is with Khalifa University of Science Technology and Research, UAE, email: harish.bhaskar@kustar.ac.ae. Lyudmila Mihaylova is with Lancaster University, UK. Email: mila.mihaylova@lancaster.ac.uk. Alin Achim is with the University of Bristol, UK, Email: alin.achim@bristol.ac.uk. We acknowledge the support of the UK MOD Data and Information Fusion DTC (Tracking Cluster project DIFDTC/CSIPC1/02) and EU COST action TU0702. Corresponding author: L. Mihaylova, mila.mihaylova@lancaster.ac.uk. Manuscript received May 2008, revised November, 2009 and March 2010. Copyright (c) 2010 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.
the background appearance method of Li et al. [11] on real
video sequences is presented in Section III. Conclusions are
summarised in Section IV.

II. THE PROPOSED CBS-SαS TECHNIQUE

In contrast with conventional BS techniques such as [6],
[24], [4] that operate at pixel level, the developed BS technique
makes use of the advantages of the cluster image representation.
Here, an image frame at time instant \( k \) is subdivided
into constituent clusters \( c_k^i \), \( (0 \leq i \leq q) \), where \( q \) is the
total number of clusters present in the image. Automatic
clustering is performed with vector quantisation and using
colour features [15].

The problem of CBS involves a decision on whether a
cluster \( c_k^i \) belongs to the background (bG) or foreground (fG)
object from the ratio of PDFs:

\[
\frac{p(bG|c_k^i)}{p(fG|c_k^i)} = \frac{p(c_k^i|bG)p(bG)}{p(c_k^i|fG)p(fG)},
\]

where, the vector \( c_k^i = (c_k^i_1, \ldots, c_k^i_m) \) characterises the \( i \)-th
cluster \( (0 \leq i \leq q) \) at time instant \( k \), containing \( f \) number of
pixels such that \( |Im|_i = |c_k^i_1, \ldots, c_k^i_f| \) is the whole image;
\( p(bG|c_k^i) \) is the PDF of the bG, evaluated using a certain
feature (e.g. colour or edges) of the cluster \( c_k^i \); \( p(fG|c_k^i) \) is the
PDF of the fG of the same cluster \( c_k^i \); \( p(c_k^i|bG) \) refers to the
PDF of the cluster feature given a model for the bG and
\( p(c_k^i|fG) \) is the appearance model of the fG object. In our
cluster BS technique the decision that a cluster belongs to a
bG is made if:

\[
p(c_k^i|bG) > \text{threshold} \left( \frac{p(c_k^i|fG)p(fG)}{p(bG)} \right).
\]

The appearance of the fG, characterised by the PDF \( p(c_k^i|fG) \), is assumed uniform. The bG model represented as
\( p(c_k^i|bG) \) is estimated from a training set \( \Re_k = \{c_k^1, \ldots, c_{k-T}\} \)
which is a rolling collection of images over a specific update
time \( T \). The time \( T \) is crucial since its update determines the
model ability to adapt to illumination changes and to handle
appearances and disappearances of objects in a scene. If the
frame rate is known, the time period \( T \) can be adapted, e.g.,
as a ratio \( T = \frac{N}{fps} \) between the number \( N \) of frames obtained
through the online process and the frame rate, fps, frames per
second. Since the threshold is a scalar, the decision in (2) is
made from the average of the distributions of all pixels within
the cluster \( c_k^i \).

A. Alpha Stable Distributions

The appeal of SαS distributions as a statistical model for
signals derives from some important theoretical and empirical
reasons [19]. Generally, there is no closed-form expression for
the PDF of SαS distributions. A convenient way of defining
them is by their characteristic function \( \varphi(c) = \exp(jpc - \gamma|c|^{\alpha}) \),
where \( \alpha \) is the characteristic exponent parameter, with
values \( 0 < \alpha \leq 2 \) that controls the heaviness of the tails of the
density function, \( \delta \) is the location parameter \((-\infty < \delta < \infty)\)
that corresponds to the mean for \( 1 < \alpha \leq 2 \), and to the median
for \( 0 < \alpha \leq 1 \) and \( \gamma \) is the dispersion parameter \((\gamma > 0)\),
which determines the spread of the density around the location
parameter. A SαS distribution, characterised by the above
three parameters is denoted by \( S(\alpha, \gamma, \delta) \). In fact, no closed-
form expressions for the general symmetric \( \alpha \) stable PDF exist,
except for the Gaussian and Cauchy members. Specifically, the
case \( \alpha = 2 \) corresponds to the Gaussian distribution
and the PDF has the form

\[
f_{\alpha=2}(\gamma, \delta; c) = \frac{1}{\sqrt{4\pi\gamma}} \exp \left\{ -\frac{(c - \delta)^2}{4\gamma} \right\}.
\]

The case \( \alpha = 1 \) corresponds to the Cauchy distribution,
for which the PDF is given by

\[
f_{\alpha=1}(\gamma, \delta; c) = \frac{\gamma}{\pi[(\gamma)^2 + (c - \delta)^2]}.\]

B. SαS Mixture Models

The PDFs \( p(c_k^i|\Re_k, bG + fG) \) of the fG and bG can be calculated as a mixture

\[
p(c_k^i|\Re_k, bG + fG) = \sum_{m=1}^{M} \pi_{m,k} \varphi(c_k^i; \alpha_k, \gamma_{m,k}),
\]

of an \( M \) component SαS PDFs \( \varphi(c_k^i; \alpha_k, \gamma_{m,k}) \) (with param-
ers \( \alpha_k \), the characteristic exponent and \( \gamma_{m,k} \), dispersion
parameter). The weighting coefficients \( \pi_{m,k} \) are calculated as
shown in the next subsection.

C. Iterative Log-Moment Estimation

The most important parameters of a SαS distribution
are the characteristic exponent \( \alpha \) and dispersion parameter
\( \gamma \). The location parameter \( \delta \) is often assumed to be
zero, i.e., the measurements are normalised with respect to
the origin. Several methods for estimating these parameters
have been introduced in [12]. In our CBS-SαS technique
the parameters of the SαS distribution are evaluated based
on the log-moment estimation method. The update of the
parameter estimates \( \hat{\alpha}_{(1,k)}, \hat{\alpha}_{(2,k)}, \ldots, \hat{\alpha}_{(M,k)} \) and \( \hat{\gamma}_{(1,k)}, \hat{\gamma}_{(2,k)}, \ldots, \hat{\gamma}_{(M,k)} \) at time instant \( k \) is performed, respectively,
from the estimates \( \hat{\alpha}_{(1,k-1)}, \hat{\alpha}_{(2,k-1)}, \ldots, \hat{\alpha}_{(M,k-1)} \) and \( \hat{\gamma}_{(1,k-1)}, \hat{\gamma}_{(2,k-1)}, \ldots, \hat{\gamma}_{(M,k-1)} \) at a previous time \( k-1 \).

According to the log moment estimation technique, if \( \Re_k \)
is a real SαS random variable, then its \( p \)-th order moment
satisfies the relation \( E(|\Re_k|^p) = E(e^{p \log|\Re_k|}) = E(e^{pV}) \),
where \(-1 < p < \alpha \) and \( V = \log|\Re_k| \) corresponds to a
log [SαS] process with \( \mu_k \) and \( \sigma_k \) representing the mean and
variance of the \( \Re_k \) samples. The estimates of the mean and
variances, \( \hat{\mu}_{k,1}, \ldots, \hat{\mu}_{k,M} \) and \( \hat{\sigma}_{k,1}^2, \ldots, \hat{\sigma}_{k,M}^2 \), respectively
of the \( \Re_k \) samples can be represented as in [23] with

\[
\hat{\mu}_{m,k+1} = \hat{\mu}_{m,k} + \frac{1}{T_k} (\hat{\sigma}_{m,k} - \hat{\mu}_{m,k}),
\]

\[
\hat{\mu}_{m,k+1} = \frac{1}{T_k} \hat{\mu}_{m,k} + \frac{1}{T_k \hat{\mu}_{m,k}} \hat{\delta}_{m,k},
\]

\[
\hat{\sigma}_{m,k+1}^2 = \hat{\sigma}_{m,k}^2 + \hat{\alpha}_{m,k} \left( \frac{1}{T_k \hat{\mu}_{m,k}} \right) \hat{\delta}_{m,k},
\]

where \( \delta_{m,k} = c_k^i - \hat{\mu}_{m,k} \), \( \mu \) denotes the transpose operation,
\( \hat{\alpha}_{m,k} \) refers to the ownership of the new cluster and defines
the closeness of this cluster to a particular SαS component,
and \( m = 1, \ldots, M \). The dispersion parameter \( \gamma_{m,k} \) at any time instant \( k \) can be updated iteratively using the following equations [12]

\[
\frac{\log \gamma_{m,k}}{\alpha_{m,k}} = \frac{k-1}{k} \left( \frac{C_e(1-\alpha_{m,k-1})+\log \gamma_{m,k-1}}{\alpha_{m,k-1}} \right)
+ \frac{\mu_{m,k}}{\kappa} + C_e \left( 1 - \frac{1}{\alpha_{m,k}} \right)
\]

and similarly the characteristic exponent \( \alpha_k \) at time instant \( k \) can be updated as follows [12]

\[
\frac{\pi^2}{6} \left( \frac{1}{\alpha_{m,k}} + \frac{1}{2} \right) = \frac{k-1}{k^2} \left( \frac{C_e(1-\alpha_{m,k-1})+\log \gamma_{m,k-1}}{\alpha_{m,k-1}} \right)
+ \frac{\pi^2(k-1)}{6k} \left( \frac{1}{\alpha_{m,k-1}} + \frac{1}{2} \right) + \frac{1}{k} \sigma_{m,k},
\]

where \( C_e = 0.57721566 \ldots \) is the Euler constant. The accuracy of this parameter estimation technique increases with the increase in sample size. The ownership of any new cluster is set to 1 for “close” components and the others are set to zero. A cluster is close to a component iff the Mahalanobis distance between the component and the cluster centre is, e.g., less than 3. If there are no “close” components, a component is generated with \( \bar{w}_{m+1,k} = 1/T_k \), with an initial mean \( \bar{\mu}_0 \) and variance \( \bar{\sigma}_0^2 \). The model presents clustering of components and the background is approximated with the proposed mixture of \( \alpha \) distributions is able to avoid over-segmentation, to cope well with the multi-modality and hence to represent in a better way the object of interest than the GMM. In the next subsection, results on video data captured with a moving camera and movements in the \( bG \) are presented.

### B. Moving Camera Video Sequence

Deep sea water video sequences are particularly challenging due to the video camera movement. The object of interest is a deep sea diver who is also in motion. There are sources of multi-modality, moving \( bG \) and the camera is moving. From Fig. 2 a) and b) is evident that the \( \alpha \) CBS reduces significantly the level of clutter. We performed additional tests on another video sequence, corresponding to an outdoor scenario (taken from [16]). The results from Fig. 3 show that the
the ratio of the number of pixels correctly identified to the number of pixels in the ground truth and precision is computed as a ratio of the number of pixels correctly identified to the number of pixels detected. In Figure 4 it can be observed that both techniques show a steady decrease in precision with increase in recall. However, the proposed CBS-\(S_\alpha S\) algorithm (blue line) displays higher levels of precision for the same values of the recall than the CBS-GMM (red dashed line). The higher the rate of precision implies a greater rate of correct classification of pixels [5].

The performance of the CBS-\(S_\alpha S\) technique was further tested over a large sequences from CAVIAR [1] and PETS [2] datasets. In addition to the recall and precision measures, the ratio \(S(A, B) = \frac{|A \cap B|}{|A \cup B|}\) [7] between the intersection of the ground truth and segmented regions over the union of the ground truth and segmented regions is used to evaluate the performance of the CBS-\(S_\alpha S\), CBS-GMM algorithms and the algorithm of Li et al. [11]. The accuracy of the segmentation process increases with increasing the values of \(S\). If \(S > 0.5\) the performance of segmentation is generally considered good and nearly perfect for values of \(S > 0.8\). The average values of precision, recall and \(S\)-ratio are shown in Table I calculated on CAVIAR data [1], from indoor environment and with the camera above the level of the walking person. According to the these results the CBS-

<table>
<thead>
<tr>
<th>Sequence Number</th>
<th>CAVIAR - Video</th>
<th>CBS - GMM</th>
<th>CBS - SAS</th>
<th>Li - GMM</th>
<th>CBS - GMM</th>
<th>CBS - SAS</th>
<th>Li - GMM</th>
<th>CBS - SAS</th>
<th>Li - GMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Browne1</td>
<td>0.42</td>
<td>0.48</td>
<td>0.48</td>
<td>0.31</td>
<td>0.47</td>
<td>0.39</td>
<td>0.35</td>
<td>0.57</td>
</tr>
<tr>
<td>2</td>
<td>Browne2</td>
<td>0.31</td>
<td>0.54</td>
<td>0.48</td>
<td>0.38</td>
<td>0.52</td>
<td>0.44</td>
<td>0.23</td>
<td>0.48</td>
</tr>
<tr>
<td>3</td>
<td>Walk1</td>
<td>0.23</td>
<td>0.47</td>
<td>0.41</td>
<td>0.41</td>
<td>0.48</td>
<td>0.42</td>
<td>0.39</td>
<td>0.61</td>
</tr>
<tr>
<td>4</td>
<td>Walk2</td>
<td>0.26</td>
<td>0.58</td>
<td>0.54</td>
<td>0.57</td>
<td>0.61</td>
<td>0.51</td>
<td>0.28</td>
<td>0.59</td>
</tr>
<tr>
<td>5</td>
<td>FightChase</td>
<td>0.32</td>
<td>0.61</td>
<td>0.53</td>
<td>0.41</td>
<td>0.53</td>
<td>0.27</td>
<td>0.31</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Table I: Precision, recall and \(S\)-ratio with CAVIAR data [1]

IV. Conclusions and Future Work

A novel approach for automatic object detection based on cluster BS with \(S_\alpha S\) distribution was introduced. The heavy tail \(S_\alpha S\) distributions allow to cope with slight movements in the background, camera shakes and clutter noise. An adaptive framework for parameter estimation is proposed that allows the model to adapt to environmental changes. A comparison of the model to its counterpart CBS-GMM model [3] is presented. Experimental results show that the CBS-\(S_\alpha S\) algorithm has efficient performance measured by precision, recall and \(S\)-ratios and outperforms both the CBS algorithms with a GMM.
and the algorithm of Li et al. [11]. The model has relatively low memory requirements and can process at the rate of 15-20 fps on a Intel Duo Core processor machine. Our future work will be focussed on BS for videos from moving cameras, with scenarios similar to the considered in [20], [17], [8].

REFERENCES