ABSTRACT

A precise object contour detection based on foreground segmentation in a video sequence is proposed. Our method employs a set of block-based Gaussian mixture models (GMM) on different block sizes, followed by foreground object contour extraction, which is based on Walsh transform (WT) domain. Every background model has different spatial and temporal correlations, and exploits WT-based feature parameters. This scheme can reduce both false positive errors and false negative errors, caused by dynamic background motions, such as global lighting changes and scenery in heavy snow. Moreover, a set of GMMs on every block size mode can be calculated by less than 10% steps, compared with a conventional pixel-wise GMM scheme. The contour detection uses the result of the foreground segmentation for extracting precise edges along the actual contour by using WT horizontal/vertical spectra feature parameters of edge blocks. In addition to simple computation, our method requires no initialization with respect to boundaries of objects by using information of foreground regions. Experimental results show robust performance of the proposed method under outdoor conditions.

Index Terms—foreground segmentation, object contour detection, Variable block size Gaussian mixture model, Walsh transform

1. INTRODUCTION

Detection of a foreground object in a video sequence is a key step in various applications, such as video surveillance, auto-cruising and video conference systems. In these scenarios, extraction of precise shapes of foreground objects is preferable for high level vision tasks, such as MPEG-4 based multimedia editing, recognition and tracking. However, in order to obtain a precise contour, one of the problems is to suppress dynamic background motions, including global lighting changes and bad weather situation.

A common approach of detecting a contour of a foreground object is to use an active contour model [1]. This approach is classified as a boundary-based contour detection which requires to model a contour by using an energy function. The objective of the approach is to get an object contour enclosing the object. The object contour is obtained by minimizing an energy function, evolving the contour toward the object region. Xu et al proposed a graph cuts based active contours to apply face contour tracking [2]. This method does not require any a priori global shape model because of the introduction of a graph cuts technique. However, the performance on the outdoor situation has not yet been reported in the method. Yilmaz et al proposed an object contour tracking method using level sets [3]. This method can be applied for tracking object contour using mobile camera at an outdoor situation, however, the algorithm requires the initialization of the boundaries of objects in the first frame.

In contrast, background subtraction techniques can be used as a region-based contour detection which is performed by using a background model. The main goal of the approaches is to label the remaining regions as foreground objects by the subtraction. One of the best formulations is to introduce a Gaussian mixture model (GMM) into every pixel in the background. Stauffer et al proposed the GMM scheme which consists of the RGB values of a pixel [4]. This scheme is suitable for handling a multi-modal background, except the sudden global lighting changes. The weakness comes from the fact that the model ignores the dependencies among neighboring pixels. In order to handle sharp changes of illumination, a hierarchical GMM was proposed in [5]. In addition to the pixel-wise GMM, this approach employs state models for the RGB values of different scales. Although this hierarchical structure quickly adapts sharp changes compared with the pixel-wise GMM, a huge amount of computations resulted due to pixel-wise processing.

Our goal is to develop the foreground segmentation method which simultaneously realizes the accuracy and the stability so as to increase video picture quality. Our system is currently assumed that the camera keeps still but the scenery includes the complex background motions. In order to realize the stability, a multi-resolutional GMM scheme
has been proposed in [6]. By merging a set of GMMs on every different block size mode, the resulting segmentation gives stable performance under outdoor situation, including sudden global lighting changes. Besides, proposed GMM scheme can be performed by less than 10% steps, compared with the pixel-wise GMM in [4]. However, the resolution of foreground segmentation is remaining issue of our method in terms of precise contour extraction. Because of the block-based operation, the resolution of segmentation becomes low. In this paper, a computationally simple object contour detection is introduced in order to detect more accurate contour of foreground objects. Our approach is based on the directional dependence of WT spectra feature parameters. In addition, because of the use of the foreground segmentation technique, our contour detection algorithm needs no initialization with respect to boundaries of objects.

The rest of the paper is organized as follows. In section 2, our foreground segmentation method is described. In section 3, proposed object contour detection based on the foreground segmentation is described. Experimental results of our method are also shown in each section. Finally, possible applications and conclusions are described in section 4 and 5.

2. FOREGROUND SEGMENTATION

2.1. Multi-resolutional GMM in Walsh Transform Domain

In this section, our foreground segmentation algorithm is briefly described. Our framework is based on a multiple block sizes GMM in Walsh transform domain processing. Feature parameters to be applied to the GMM is defined by Walsh spectra feature parameters. The use of GMMs on texture related parameters in addition to the luminance parameter in the transform domain is expected to reduce false negative errors, compared with the pixel-wise, luminance-based GMM. Because of the introduction of different block-based GMM, false positive errors can be reduced.

A flow chart of our foreground segmentation algorithm is shown in Fig. 1. Our background modeling scheme has different block size modes. The block sizes are set to \( N = 2^n \), where \( n \) is a positive integer \( n \in \{2, 6\} \). The tentative segmentation results on every block are obtained from OR operations of luminance and texture-based decisions. All of the results are merged by AND operation to give final segmentation result in the frame.

The luminance and texture-based GMM approach employed on every block in Fig. 1 is carried out in the following way. On each block size mode, two feature parameters are determined by Walsh transform of the luminance components as follows:

\[
X_{DC,t} = f_0
\]

\[
X_{AC,t} = [f_v, f_h, f_d]^T,
\]

where,

\[
f_0 = \text{WAL}(0, 0)
\]

\[
f_v = \sum_{i=0}^{3} i^2 \times |\text{WAL}(i, 0)|
\]

\[
f_h = \sum_{j=0}^{3} j^2 \times |\text{WAL}(0, j)|
\]

\[
f_d = \sum_{i=1}^{3} \sum_{j=1}^{4-i} (i^2 + j^2) \times |\text{WAL}(i, j)|.
\]

\(|\text{WAL}(i, j)|\) is an absolute value of the Walsh coefficient of coordinate \((i, j)\), \(f_0\) is the DC coefficient, and \(f_v, f_h\) and \(f_d\) are weighted sums of AC coefficients in vertical, horizontal and diagonal directions, respectively. This is because every coefficient has strong spatial correlation within the set.

Let \(X_t\) be a feature parameter at time \(t\), the GMM scheme assumes the probability of the observing parameter by a
mixture of $K$ Gaussian distributions in the following way.

$$P(X_t) = \sum_{i=1}^{K} \omega_{i,t} \times \mathcal{N}(\mu_{i,t}, \Sigma_{i,t}),$$  \hspace{1cm} (4)

where $\omega_{i,t}$ is the weight of the $i$-th Gaussian, $\mathcal{N}(\mu_{i,t}, \Sigma_{i,t})$ indicates the Gaussian probability density function of $i$-th component, having the mean $\mu_{i,t}$ and the covariance matrix $\Sigma_{i,t}$. In our scheme, $K = 3$ is used in every block size mode. When Eq. (2) is applied to the GMM, the covariance matrix is determined as the following form,

$$\Sigma_{DC,i,t} = \sigma_{0,i,t}^2$$
$$\Sigma_{AC,i,t} = \text{diag} \left[ \sigma_{v,i,t}^2, \sigma_{h,i,t}^2, \sigma_{d,i,t}^2 \right],$$  \hspace{1cm} (5)

where $\text{diag}[\cdot]$ indicates the diagonal matrix. This assumes that the three elements of Eq.(2) are independent and have different variances.

The GMM on each block size mode is composed of one or more Gaussians to the background model, and the foreground is modeled by remaining Gaussians. Therefore, $B$ out of $K$ distributions are considered as the background model, based on the weight of each Gaussian and its standard deviation. If $X_t$ satisfies

$$|X_t - \mu_{i,t-1}| < T,$$  \hspace{1cm} (6)

and the Gaussian with $\mu_{i,t-1}$ was included in the $B$ distributions, then $X_t$ is classified as the background. $T$ is a threshold to judge whether the current state of $X_t$ is matched in the distribution of the $i$-th Gaussian. This threshold $T$ is defined for DC and AC decisions respectively:

$$T_{DC,N} = \delta_{DC,N} \times \mu_{DC} + 2.5 \times \sigma_{DC}$$
$$T_{AC,N} = \delta_{AC,N} \times \mu_{AC} + 2.5 \times [\sigma_v, \sigma_h, \sigma_d]^T,$$  \hspace{1cm} (7)

where $\delta_{DC,N}$ and $\delta_{AC,N}$ are parameters which may have different values on each block size $N$. When $X_t$ satisfies Eq. (6), the update scheme based on the Ref. [7] is applied to the $\mu_{i,t-1}, \Sigma_{i,t-1}$ and $\omega_{i,t-1}$ of the distribution.

2.2. Experimental Results on Foreground Segmentation

Three different foreground segmentation algorithms are compared on three test sequences, shown in Fig. 2. Test sequences were taken in outdoor situation with a CMOS sensor camera, recording to HDV format. The size of the sequences was stretched to $1920 \times 1080$ pixels. Ground truths shown in Fig. 2 (b) are manually labeled by pixel. Our test sequences have the following features: The test sequence 1 contains that two person are crossing each other and their brightness is dark. The test sequence 2 includes global lighting changes, due to the covered/uncovered sun by a cloud. In addition, the leaves around the pond and waves at the surface of the pond causes the complicated motions in the background. The test sequence 3 was taken in the heavy snow situation. The snow falls toward a diagonal direction.

Our method is compared with the pixel-wise GMM scheme and a Discrete Cosine Transform (DCT) based single Gaussian model scheme proposed [4, 7]. The method proposed in [7] is similar to our method in terms of the employment of transform domain processing, however, the method uses a single Gaussian model instead of the GMM for reducing the computation amount. In our experiment, false positive ratio (FPR) and false negative ratio (FNR) are used as the metrics of foreground segmentation performance. FPR and FNR are defined as follows.

$$\text{FPR} = \frac{\text{the number of background pixels wrongly classified}}{\text{the number of background pixels in the ground truth}}$$
$$\text{FNR} = \frac{\text{the number of foreground pixels wrongly classified}}{\text{the number of foreground pixels in the ground truth}}$$
Fig. 3. Evaluation of the foreground segmentation algorithms by means of False Positive Ratio (FPR) and False Negative Ratio (FNR) for selected test frames on the test sequences.

The results of performance evaluation using FPR and FNR are summarized in Fig. 3. The parameters on the algorithms are set to the suggested values which are described in the references [4, 7, 8].

The pixel-wise GMM method presents high FPR on test seq. 2 in Fig. 3 (a), due to the weak nature for the global lighting changes. The method produces many false negatives within an object on test seq. 1 in Fig. 3 (b), due to dark pictures. These error comes from the pixel-level and the spatially independent processing. The adoption of spatial correlation makes algorithms robust to the global lighting changes. Our approach and the DCT-based approach shows smaller FPR than that of the pixel-wise GMM approach. The use of the single Gaussian model may lead to false errors. For example, a single foreground object is separated into a few foreground blobs on test seq. 1, shown in Fig. 2. Moreover, the single Gaussian model can not deal with the bi-modal background motion, such as heavy snow on seq. 3 in Fig. 2.

Our method achieves the lowest FNR on the average. This is because the block-based GMM with different block sizes can perform a multiple threshold scheme for employing lower thresholds in smaller block sizes. In terms of FPR, because our method has different spatial correlation, false positive error caused by background motions can be reduced. However, FPR of our method is slightly higher than that of the DCT-based method. The result is caused by a small number of foreground detection in the DCT-based method. On the other hand, the block-wise detection around the edges of an object in our method is also the factor to produce the error. Therefore, in order to make our method into a more accurate system, an object contour detection method is introduced.

3. OBJECT CONTOUR DETECTION

3.1. Walsh spectra feature parameters and contour decomposition

The proposed contour detection is based on the foreground segmentation. The method exploits Walsh spectra feature parameters, applied to our GMM scheme, shown in Eq. (2). The blocks on contoured part of the detected foreground blob are halved in a repetitive manner by using directional dependence in vertical/horizontal components of Walsh spectra.

A flow chart of our object contour detection algorithm is shown in Fig. 4. First, the block-wise contour of foreground blob is extracted, which is based on the fact that our foreground segmentation does not produce separated blobs in terms of a single object. Second, Walsh feature parameters $f_v$ and $f_h$ are compared only in the edge blocks of $N \times N$, where $N$ is the smallest size applied to foreground segmentation. If the sum of vertical Walsh spectra components is larger than horizontal class, the upper/lower side of the block apart from the centroid is removed, resulting $N \times N/2$ blocks.
3.2. Experimental Results on Contour Detection

The proposed contour detection is evaluated by comparing the results of our foreground segmentation on the test sequences, shown in Figs. 5 (a) and (b). In order to evaluate the approaches, the ground truth is first depicted manually, shown in Fig. 5 (c).

The results shown in Fig. 5 (d) have irregular shapes, due to the block-wise detection. On the other hand, by using the contour detection, an object with further decomposition gives us natural extraction from video on the 53 frame of the test seq. 1, shown in Fig. 5 (e). Moreover, the result on the 26 frame in the test seq. 2 shows that jagged shape of detected contour around the legs of person can be reduced. In case of the test seq. 3, a car and its shadow are detected, due to the textured background. Because the shadow region is detected as the foreground at the stage of foreground segmentation, a region existing ideal contour cannot be covered by the edge blocks. Therefore, the exact contour of the car cannot extracted on the frame, shown in the third row in Fig. 5.

A quantitative evaluation is summarized in Fig. 6. In addition to FPR and FNR, a sum of false positive and false negative is also shown as a total error. Average FPR and
The proposed contour detection decreases FPR, caused by the block-wise processing, while FNR is increased, due to excessive decomposition of the edge blocks. However, up to 29.7% improvement has been observed in the test seq. 1 in terms of the total error. Average rates of the improvement are 20.2, 5.88 and −1.46% for the test seq. 1, 2 and 3, respectively.

### 4. POSSIBLE APPLICATION OF SPATIO-TEMPORAL GAUSSIAN MIXTURE MODEL

In this section, possible application of our method is described. Our foreground segmentation and contour detection methods are based on the spatio-temporal GMM. Therefore, our spatio-temporal GMM enables applying to the camera movement, while the pixel-wise GMM is only applicable to a fixed camera. This is because the pixel-wise GMM requires the accurate alignment of pixel position. On the other hand, the spatio-temporal GMM makes this requirement a little bit loose, due to the employment of the spatial correlation. Therefore, our method can be applied to pan/tilt/zoom camera. An trial of our foreground segmentation algorithm to videos from pan/tilt/zoom cameras is now under study.

As an application of our algorithm in terms of video/picture quality improvement, background separation coding is considered for low bit-rate, clear video transmission for the purpose of monitoring several different surveillance results at a center. The background separation coding with color enhancement enables reducing the required bit-rate by half, keeping 8 Mbps H.264 picture quality. The background video is first scaled down to NTSC with 15 frames per second and encoded at around 300 kbps by a H.264 encoder. On the contrary, the foreground segments are first enhanced the picture quality by a multi-scale Retinex technique and filled them into a series of blue pictures. As a H.264 encoder has the intra-frame prediction mode, around 300 kbps bit-rate is resulted in case of the Variable Bit-Rate mode of 3.5 Mbps in average and maximum rate, when no foreground is detected. Generally, the Retinex technique generates a lot of high frequency components and this fact leads to around allowable maximum bit-rate on a non-separation coding and also degrades image quality. However, the method of the background separation coding enables clear images under 4 Mbps.

### 5. CONCLUSIONS

The precise object contour detection using the foreground segmentation technique has been proposed. Our method employs the multi-resolutional Gaussian mixture model, followed by the Walsh transform-based contour detection which is based on the directional dependence of Walsh spectra. By using the spatio-temporal GMM, stable performance is achieved under different conditions of the background, such as the global lighting changes and heavy snow situation. In addition to the method, the contour detection results in the accurate contour of an object by decomposing the block-wise detection around the edges. Experimental results show that the use of contour detection gives up to 29.7% improvement in terms of false positives and false negatives in outdoor situation. To cope with the shadow of foreground object and textured background such a bad weather situation is future tasks.

### 6. REFERENCES


