Detection and Decomposition of Foreground Target from Image Sequence

K. L. Chan
Department of Electronic Engineering, City University of Hong Kong
83 Tat Chee Avenue, Kowloon, Hong Kong
itklchan@cityu.edu.hk

Abstract

A surveillance video contains two components – the background scene and the foreground targets. Foreground detection is difficult when there are illumination changes and background motions in the scene. We propose a two-step background subtraction framework for foreground detection. The background is modeled as block-based color Gaussian mixture model. In the first step of background subtraction, the current image frame is compared with the background model via a spatial similarity measure. The potential targets are separated from most of the background pixels. In the second step, if a potential target is sufficiently large, the enclosing block is compared with the background model again in order to obtain a refined shape of the foreground. Complex target shape may exhibit multiple motions. Decomposition of the foreground region into meaningful parts is essential for the recognition of the activity. We adopt the morphological shape decomposition algorithm to decompose each foreground region. The method is enhanced by considering color cue in the decomposition process. We test the foreground detection and decomposition methods on a swimming image sequence.

1. Introduction

Automatic visual surveillance systems for human motion monitoring typically consist of the detection and tracking of human along the image sequence, and then the inference of the motion. The human subject is often considered as the foreground. Its detection can therefore be achieved via the background subtraction process. Many interesting foreground targets may exhibit multiple motions in the image sequence. For instance, a walking human has translational motion in the trunk, whilst the hands and legs are swinging (translation plus rotation). Decomposition of the foreground region into meaningful parts and then tracking of the body parts are essential for the recognition of the activity. In this paper, we propose a new background subtraction method for human target detection. We also develop a foreground decomposition technique that can segment the human target region into trunk, hands and legs by using both shape and color cues.

One common assumption in background subtraction is that the background is stationary or changes slowly. However, pixels may exhibit multiple background colors due to repetitive motions and illumination changes. Some methods have been proposed that can tackle background movements. For instance, Stauffer *et al.* [1] claim the pixelwise mixture of Gaussians can deal with repetitive motions of scene elements. However, the results in [2] and [3] indicate that the pixelwise mixture of Gaussians is not effective in modeling dynamic background such as swaying trees, waving water, etc. Eng *et al.* [4] partition the background frame into blocks and model each block of background colors using hierarchical k-means clustering. They implement a spatial searching process to detect the displaced background colors. Our method also models the background scene by a block-based scheme since it is a better approach to tackle background motions than the pixelwise scheme. We define the background color similarity measure which is different from Eng *et al.* [4]. We implement a two-step approach to detect and subtract background colors in a coarse-to-fine manner.

We are inspired by the research works in dynamic texture detection. Doretto *et al.* [5] consider the image sequence of moving scene as dynamic texture. They propose a method to model, recognize and synthesize the visual signals close to moving scene. Chetverikov *et al.* [6] address two related problems: detect regions of dynamic texture and detect targets in a dynamic texture. Dynamic texture is modeled as optical flow residual. We consider the problem of background subtraction as a searching problem in the background model such that there are colors in the background model very similar to the colors in the current video frame. In the first step of background subtraction, the dynamic and static background colors are detected by a spatial similarity measure. Finally in the second step, the initial foreground regions are refined by the subtraction of background colors via a close-range similarity measure.

Xu [7] proposes a morphological shape segmentation algorithm. Liu *et al.* [8] propose a convex shape decomposition method which minimizes the total cost of decomposition under concavity constraints. For visual surveillance application, convexity of the segmented parts is not the only requirement. We want the decomposed shapes correlate with the body parts. Therefore we adopt and modify Xu's method [7] by considering color cue in the decomposition process.

2. Foreground Detection

Our foreground detection method, as shown in Figure 1, is composed of background modeling and dynamic background subtraction. Background frame is generated from the original image sequence by vector median

filtering. The background scene is modeled by block-based k-means clusters of the background frame. We consider the problem of background subtraction as a searching problem in the background model such that there are colors in the background model very similar to the colors in the current image frame. With the use of a stationary camera in the acquisition of the video, it is reasonable to assume that the background scene does not move over a long distance. Therefore, a dynamic background can be found in nearby regions of the background model.

Due to complex background scene and wide ranging foreground colors, there are two problems in the first step result. First, some background colors cannot find a match in the background model due to large change of colors resulted from motion or illumination change. Second, some foreground colors are wrongly regarded as background. In order to obtain a refined foreground region, we implement the second step of background subtraction. To reject the false positive errors, we examine the foregrounds detected in the first step. If the foreground pixels can cluster to form a sufficiently large region, the image space enclosing that potential foreground is allowed to proceed to the second step of background subtraction. To reject the false negative errors, we limit the search space in the second step of background subtraction. Each pixel of the potential foreground image space is compared with the background model at close proximity. This is to avoid the foreground color to find a match with a neighboring background model. The algorithm of our foreground detection method is shown in the pseudo-code below.

Partition current image frame into blocks
Step 1:
For each pixel
  Calculate similarity of pixel with neighboring background model
    If max(similarity) < DT_{far}
      Label pixel as potential foreground
    Else
      Label pixel as background
Step 2:
For each block of pixels
  If size of potential target > N_{target}
    For each pixel
      Calculate similarity of pixel with background model at close range
      If max(similarity) < DT_{near}
        Label pixel as foreground
      Else
        Label pixel as background

Each image frame is partitioned into n_1 x n_2 nonoverlapping blocks $B_{a,b}$, where $1 \le a \le n_1$ and $1 \le b \le n_2$. The block size is the same as in the background/foreground modeling. In step 1, pixels are classified as potential target or background. For each pixel p in $B_{a,b}$, background models of the enclosing block and neighboring blocks are used in the similarity measure

$$S_{c,m,n,C} = \frac{1}{d}\sum_{i=1}^{d}\exp\left(-\left(\frac{c^i - \mu^i_{m,n,C}}{\mu^i_{m,n,C}}\right)^2\right)$$

where c^i is the i^{th} color component, and $\mu^i_{m,n,C}$ is the mean value of the i^{th} component of a background color C in block $B_{m,n}$, m = a-1:a+1, n = b-1:b+1. If the pixel belongs to background, at least one background model is close and the corresponding similarity measure is large (near to 1). If the pixel is not a background, no background models are close and all similarity measures are low. Thresholding of the similarity measure is governed by the parameter DT_{far}. If $max(S_{c,m,n,C}, \forall m,n,C) < DT_{far}$, c is labeled as a potential target pixel. Otherwise it is a background pixel.

In step 2, if the number of potential target pixels in a block is sufficiently large, a true target may be present. Otherwise, this block belongs to background. This filtering process aims to remove small and scattered foreground regions. Each pixel of the target containing block is compared with the background model of that block by a close range similarity measure

$$S_{c,C} = \frac{1}{d}\sum_{i=1}^{d}\exp\left(-\left(\frac{c^i - \mu^i_{a,b,C}}{\mu^i_{a,b,C}}\right)^2\right)$$

where $\mu^i_{a,b,C}$ is the mean value of the i^{th} component of a background color C in the enclosing block $B_{a,b}$. If the pixel belongs to background, at least one background model is close and the corresponding similarity measure is large (near to 1). If the pixel is not a background, no background models are close and all similarity measures are low. Thresholding of the similarity measure is governed by the parameter DT_{near}. If $max(S_{c,C}, \forall C) < DT_{near}$, c is labeled as a target pixel. Otherwise it is a background pixel.

3. Part Segmentation

In Xu's method [7], a binary shape is eroded continuously until a set of skeleton points is obtained. Starting from the highest-order skeleton points, connected components (skeleton segments) are identified. Each skeleton segment can merge with a neighboring shape segment to form a larger part as far as the new shape segment is roughly convex. Otherwise that skeleton segment forms its own shape segment. All shape segments will be dilated. The whole process will be repeated until the lowest-order skeleton points are processed. Xu's method does not utilize image properties in the decomposition process. Also, as the author has mentioned, the method may not work well in segmenting elongated parts. As for the human target, we find that Xu's method can often merge the trunk and limbs into a single shape segment. For visual surveillance applications, we need to decompose each human foreground into trunk and limbs. We therefore propose a modified Xu's method. First, we control the original Xu's method to produce more shape segments. Then, we add a shape segment merging step. Neighboring shape segments (SH_a and SH_b) can be merged if they have similar colors. The color similarity

is measured by $\left|m^i(SH_a) - m^i(SH_b)\right|$, where m^i is the mean value of the i^{th} color component. If all color similarity measures are below the pre-defined threshold value, SH_a and SH_b will be merged to form a single shape segment. Before the decomposition, each foreground detection frame is pre-processed by morphological closing and opening operations to refine the foreground region, and eliminate scattered false positive errors. The algorithm of our foreground decomposition method is shown in the pseudo-code below with the modifications highlighted.

Pre-process foreground detection result
Generate skeleton points
Start from highest-order skeleton points
Repeat
  Identify skeleton segments
    For each skeleton segment
      For each existing shape segment
        If skeleton segment + shape segment is convex
          Merge skeleton segment and shape segment
      If there is no merging
        Form a new shape segment
  Dilate all shape segments
  Decrement order
For each shape segment
  Find neighboring shape segment
  If similar in color
    Merge the two shape segments

4. Result

We test the foreground detection and decomposition methods on a swimming image sequence. The swimmer at the center swims quickly towards the camera in butterfly. The swimmer on the right swims slowly in freestyle towards the camera. The top row of Figure 2 shows some original image frames of the sequence. We compare our method with another foreground detection method codebook (CB) [9]. As shown in Figure 2, CB produces more false positive errors. Our method can detect a fairly good shape of both swimmers. Table 1 shows the average values of the quantitative measures from frames 480 to 518. As compared with CB, our method achieves the higher Precision and F-Measure values.

The right swimmer is very small and so in most image frames there is only one shape segment. The decomposed body parts of the center swimmer are displayed in different grey levels. For comparison, the part segmentation results obtained by Xu's method are also shown. We relax the convexity constraint in Xu's method so that the number of shape segments generated is approximately the same as our method. As shown in Figure 2, Xu's method often merges the trunk and limbs into a single shape segment.

Table 1. Numeric results of two background subtraction methods.

	CB	Our method
Recall	**0.7198**	0.6347
Precision	0.2637	**0.8033**
F-Measure	0.3708	**0.7064**

5. Conclusion

We develop a method for the detection of foreground in a scene containing vigorous changes and motions via a two-step background subtraction. In the first step, static and dynamic background pixels are rejected while the potential targets are identified. In the second step, each sufficiently large target is checked with the background model proximally in order to obtain a refined shape of the true foreground. We adopt and modify a morphological shape decomposition algorithm to decompose each foreground region into meaningful parts. The method is enhanced by considering color cue in the decomposition process.

References

[1] C. Stauffer, W.E.L. Grimson. "Learning patterns of activity using real-time tracking," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 8, 747-757, 2000.

[2] Y. Zha, D. Bi, Y. Yang. "Learning complex background by multi-scale discriminative model," Pattern Recognition Letters, Vol. 30, 1003-1014, 2009.

[3] F. El Baf, T. Bouwmans, B. Vachon. "Type-2 fuzzy mixture of Gaussians model: application to background modeling," Proceedings of International Symposium on Visual Computing, LNCS 5358, Part I, 772-781, 2008.

[4] H.-L. Eng, J. Wang, A.H. Kam, W.-Y. Yau. "Robust human detection within a highly dynamic aquatic environment in real time," IEEE Transactions on Image Processing, Vol. 15, No. 6, 1583-1600, 2006.

[5] G. Doretto, A. Chiuso, Y.N. Wu, S. Soatto. "Dynamic textures," International Journal of Computer Vision, Vol. 51, No. 2, 91-109, 2003.

[6] D. Chetverikov, S. Fazekas, M. Haindl. "Dynamic texture as foreground and background," Machine Vision and Applications, Vol. 22, No. 5, 741-750, 2011.

[7] J. Xu. "Morphological decomposition of 2-D binary shapes into simpler shape parts," Pattern Recognition Letters, Vol. 17, 759-769, 1996.

[8] H. Liu, W. Liu, L.J. Latecki. "Convex shape decomposition," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 97-104, 2010.

[9] K. Kim, T.H. Chalidabhongse, D. Harwood, L.S. Davis. "Real-time foreground-background segmentation using codebook model," Real-Time Imaging, Vol. 11, 172-185, 2005.

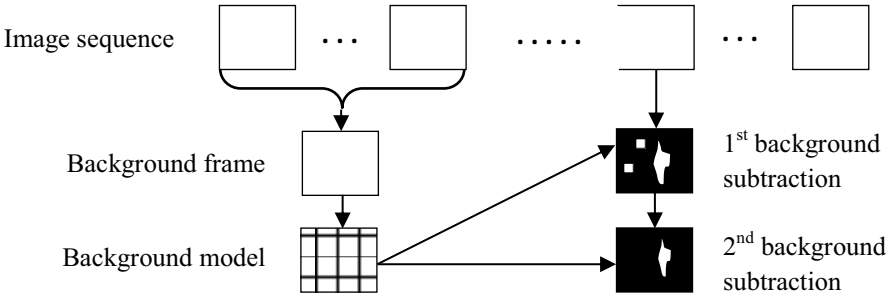

Figure 1. Overview of foreground detection.

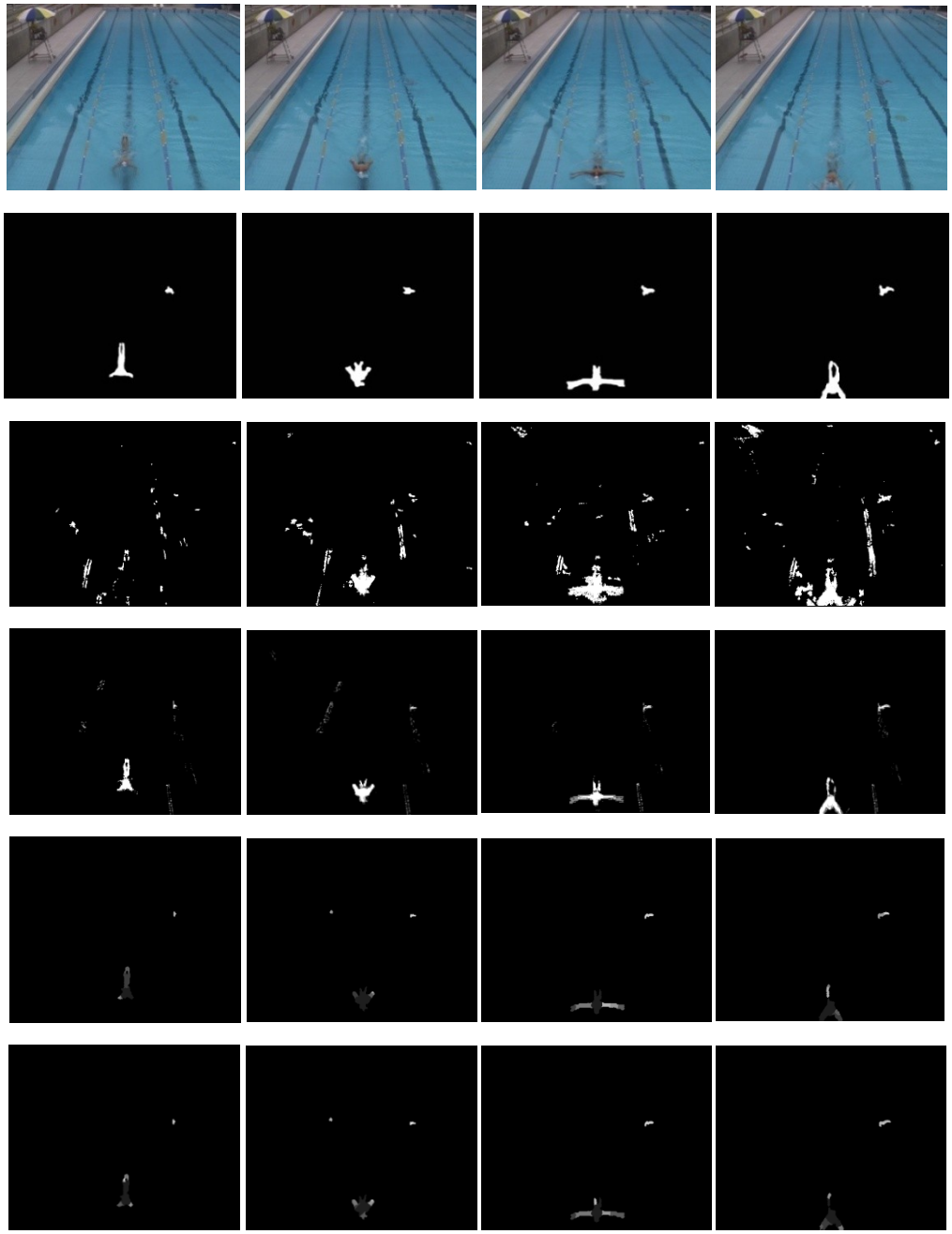

Figure 2. Results of swimming image sequence: original image frames (top row), corresponding ground truths (second row), foreground obtained by CB (third row), foreground obtained by our method (fourth row), part segmentation by Xu's method (fifth row), part segmentation by our method (bottom row).