Panoramic Background Generation using Mean-Shift 
in Moving Camera Environment

Sang-Hyun Cho  
Dept. of Computer Engineering,  
The Catholic University of Korea  
Bucheon, Korea  
cshgreat@catholic.ac.kr

Hang-Bong Kang  
Dept. of Digital Media,  
The Catholic University of Korea  
Bucheon, Korea  
hbkang@catholic.ac

Abstract— In this paper, we present a novel method to generate a panoramic background image from a sequence of moving camera images. Our method is based on non-parametric statistics and robust model estimation using mean-shift. Each frame is projected to panoramic space by homography matrix which is calculated by SURF(Speeded Up Robust Features) matching. To generate reliable panoramic background image, we use mean-shift. Our proposed method allows the generation of very clear panoramic backgrounds in moving camera environment. Experimental results from real sequences demonstrate the effectiveness of our method.

Keyword: background generation, mean-shift, background modeling, object detection, object tracking

I. INTRODUCTION

Object detection and tracking are very important in various computer vision applications such as video surveillance systems and monitoring system. In many surveillance or tracking systems, a background image with no moving objects is used as the reference information for an event analysis in detail. Moreover, one of the most adopted approaches for moving objects detection is background subtraction. The principle of background subtraction is to compare each present frame in the image sequence to the reference background image. So, it is important to build and maintain the background model of a scene.

Many background generation methods were proposed but the fundamental problems for accurate object detection or tracking are still far from being completely solved. Most of them are valid only in static camera environment. In static camera environment, the average image of all frames is widely used as a background image due to its simplicity. However, this method may not be efficient since it must use the pixel information of all frames for calculation. And it also has a blurring problem if some used frames are containing any moving objects. Recently, because of these problems, the kernel density estimation based background generation method like mean-shift is widely used.

In this paper, we propose a novel method to generate panoramic background model using mean-shift in moving camera environment. Our method uses panoramic technique to obtain a panoramic image from the image sequence. Each frame is projected into common panoramic space and reliable background mode is detected by mean-shift.

The rest of this paper is organized as follows. Section 2 explains related works. Section 3 presents our background generation method. Section 4 discusses our experimental results.

II. RELATED WORK

The background modeling method can be classified into parametric or non-parametric. The parametric method assumes that the probability density function of pixels is represented by the specific model such as Gaussian or mixture of Gaussian and that their parameters are estimated based on the sample data [1-3]. Stauffer and Grimson [4] dealt with motion segmentation problem using adaptive background subtraction. They modeled each pixel into Gaussian mixture probability density function and used online approximation in order to update the model. In the W4 surveillance system[5], both pixel-based update and object-based update are employed for background maintenance. Pixel-based update is designed to reflect the illumination change in the scene, and the object-based update is used to reflect the real physical movement in the background. W. Wren et al. [1] suggested Pfind using a single Gaussian model. Though these methods were among the first to principally model the uncertainty of each pixel color, it was quickly found that all these specific models were ill-suited to most outdoor situations since repetitive object motion, shadows or reflectance often caused multiple pixel colors to belong to the background at each pixel.

The non-parametric method has an advantage for arbitrary data because it doesn’t need a specific model unlike the parametric method. The simplest non-parametric
background generation method is to calculate the background pixel value at the mean or the median value of pixel value distribution along the time axis. Taking the median value provides better results than the mean. However, such methods fail to accurately reflect the real background of the scene and is not correctly adapted to changes in the scene. Elgammal et al. [6] estimated the probability density function of each pixel value using the kernel density estimation. Recently, Yazhou Liu et al. [7] suggested non-parametric background generation method using mean-shift to detect the most reliable background model (MRBM). D. Sidibé et al. [8] suggested non-parametric statistics and mode estimation using quasi-continuous histograms (QCH) framework. Kang et al. [9] proposed an adaptive background generation method which used a geometric transform-based mosaic method. Sinha and Pollefeys [10] used high-resolution images to stitch a panorama so the panorama can give a high zoom scene and detailed information. Reilly et al. [11] used a registered background model which consists of median background image for each camera to remove global camera motion.

This paper presents a method belonging to the non-parametric category and shares similar ideas with Liu’s and D. Sidibé works [7,8]. Our proposed method is to extend these methods into the moving camera environment. Our method generates a panoramic background image by homography based projection and mean-shift. Each frame is projected into the panoramic space which consists of time axis and image coordinate axis. For projected frame image, we perform a mean-shift procedure to detect most reliable background mode in panoramic space.

### III. Panoramic Background Generation Using Mean-Shift

#### A. overview

Figure 1 shows the overview of our method. First, we extract feature points and calculate their descriptors. Since video frame is sequential, we need not to consider the matching order. So, we compute the homography between adjacent frames. Each frame is projected into panoramic space which consists of time axis and image coordinate axis. For projected frame image, we perform a mean-shift procedure to detect most reliable background mode in panoramic space.

#### B. Projection on panoramic space

The first step in the panoramic projection is to extract and match SURF features between adjacent frames. SURF is a robust image detector and descriptor, first presented by Herbert Bay et al. in [12], which can be used in computer vision tasks like object recognition or 3D reconstruction. SURF is based on sums of approximated 2D Haar wavelet responses and makes an efficient use of integral images. As basic image features it uses a Haar wavelet approximation of the determinant of Hessian blob detector.

Figure 2 shows the result of extraction of SURF feature points and matching between adjacent frames. To generate panoramic image from the sequence, we use image stitching method proposed by M. Brown [13, 14].

Assuming that the camera rotates about its optical center, the group of transformation the images may undergo is a special group of homographies. We parameterize each camera by a rotation vector \( \theta = [\theta_x, \theta_y, \theta_z] \) and focal length \( f \). So, we define pairwise homographies as follows

\[
\tilde{u}_i = H_{ij} \hat{u}_j
\]

where \( H_{ij} = K_{ij} R_i R_j^T K_{ij}^{-1} \),

\[
K_i = \begin{bmatrix} f_i & 0 & 0 \\ 0 & f_i & 0 \\ 0 & 0 & 1 \end{bmatrix}
\]
Figure 2: Result of SURF extraction and matching between adjust frames in test sequence.

Figure 3: Result of image alignment of adjust frame in test sequence.
\[ R = e^{[\theta]} \]
\[
\begin{bmatrix}
\theta_1 \\
\theta_2 \\
\theta_3
\end{bmatrix}
\]
and \( \mathbf{u}_1, \mathbf{u}_2 \) are the homogeneous image positions.

We use RANSAC(RANdom SAmple Consensus) to select a set of inliers that are compatible with a homography between adjacent frames. RANSAC is a robust estimation procedure that uses a minimal set of randomly sampled correspondences to estimate image transformation parameters, and finds a solution that has the best consensus with the data.

In the case of panoramas, we select sets of \( r=4 \) feature correspondences and compute the homography \( H \) between them using the DLT(Direct Linear Transformation) method.

To render a high quality panoramic image, we use automatic panorama straightening and multi-band blending. Automatic panorama straightening is used to correct wavy effect of panoramic image which applies a global rotation such as “up-vector” that is normal to the plane containing the camera center and the horizon. Multi-band blending is the blending strategy for unmodelled effect such as vignetting, radial distortion and so on. More detail is in [13,14]. Figure 3 shows the result of image alignment of adjacent frames.

C. Background Generation using mean-shift

To generate the first phase of background, we assume that background pixel value is the most frequent pixel value in the input image sequence. We choose the background pixel value as the stable one for each pixel. Since the real background pixel usually has high density, we use mean-shift algorithm to detect the location of the high density. Mean-shift is an efficient method for mode search proposed by Fukunaga [15] and has been proven by Comaniciu and Meer [16, 17, 18] to have a smooth trajectory property and convergence. Background generation method using Mean-shift is as follows

- Sample selection : We select a sample \( S = \{x_i\}, i=1,\ldots,n \) for each pixel, where \( x_i \) s are the intensity values of the pixel along the time axis and \( n \) is sample size.

- Local mean points’ selection: To reduce the computational load, a set of \( m (<n) \) local mean points is used. We denote this set by \( L = \{l_i\}, i=1,\ldots,l \) where \( m \) is mean of sample set \( S \).

- Mean-shift procedure :

For intensity value \( p \) of each location of pixel

1. Initialize the starting point of mean shift procedure as : \( y_i = p_i \)
2. Apply the mean shift procedure until convergence

\[
y_{t+1} = \frac{\sum_{i=1}^{n} x_i g\left(\frac{y_t - x_i}{h}\right)}{\sum_{i=1}^{n} g\left(\frac{y_t - x_i}{h}\right)}
\]  

(2) where \( g(x) \) is profile of kernel function

3. If \( \|y_{t+1} - y_t\| < \varepsilon \) (threshold) then set to intensity value of generated background.

- Derive the candidate background mode.

Since convergence points by mean-shift are very close or even identical to each other, these convergence points can be clustered into \( m \) (\( < l \)) classes. We set \( M = \{m_i, w_i\}, i=1,\ldots,m \) where \( m_i \) is the intensity value and \( w_i \) is the weight of each cluster center.

- Obtain the background mode:

\( \hat{M} = m_{\hat{i}} \), where \( \hat{i} = \arg \max_i \{w_i\} \) and \( \hat{M} \) is the background mode.

For efficient calculation, we use the initial value of each pixel as panoramic image calculated by previous step.

IV. EXPERIMENTAL RESULTS

Our proposed method is implemented using Microsoft Visual C++ and OpenCV library. We used Epanechnikov kernel function in mean-shift step. For efficient calculation, we use sampled frames to generate a background image.

Figure 4 shows test result of 50 frame test sequence 1. The video size is 360 × 240 and the frame rate was 15 fps. The RGB color space was taken as feature space.
Figure 4: background generation result of test sequence 1
(a) frame #1  (b) frame #15  (c) frame #25  (d) frame #45
(e) average panoramic image  (f) proposed method

Figure 5: background generation result of test sequence 2
(a) frame #1  (b) frame #180  (c) frame #320  (d) frame #400
(e) average panoramic image  (f) proposed method
Figure 6: background generation result of test sequence 3
(a) frame #14  (b) frame #109  (c) frame #180  (d) frame #229
(e) average panoramic image  (f) proposed method
Figure 4 (a)–(d) show some selected frame images and Figure 4(e) shows the average panoramic background generation result by the average of overall frame images. Figure 4(f) shows the result of our proposed method. As shown in figure 4, a person is moving and camera is moving from left to right. Thus, average panoramic background image is blurred and foreground is faintly remained. However, our proposed method shows specific shapes and non-blurred background compared to the average panoramic background image since the mode of the intensity distribution information is very useful in distinguishing the moving object.

Figure 5 shows 450 frame test sequence 2. The video size is 360 × 240 and the frame rate was 15 fps. The RGB color space was taken as feature space. Figure 5 (a)–(d) show some selected frame images and Figure 5(e) shows the average panoramic background generation result by the average of overall frame images. Figure 5(f) shows the result of the proposed method.

Figure 6 shows 1022 frame test sequence 3. The video size is 360 × 240 and the frame rate was 15 fps. Figure 6 (a)–(d) show some selected frame images and Figure 6(e) shows the average panoramic background generation result by the average of overall frame images. Figure 6(f) shows the result of the proposed method.

As shown in figure 5 and 6, in this test sequence, camera is moving freely. Similarly, average panoramic background image for test sequence is quite blurred and ghost objects are remained. However, our proposed method shows a reasonable background image.

V. CONCLUSION

In this paper, we present a novel method for generating a panoramic background image from a sequence of moving camera images. Our method generates a panoramic background image by homography based projection and mean-shift. Each frame image is projected into the panoramic space by homography matrix which is calculated by SURF matching. To generate reliable panoramic background image, we use mean-shift. Experimental results show that our method is effective and robust.

Our proposed method, however, has some drawbacks. The generated background result is not good when containing slowly moving object in the scene. Also, for robust background generation, it is necessary to change current pixel based method to a region based method by fusing spatial context.

Acknowledgment

This research is supported by Defense Acquisition Program Administration and Agency for Defense Development under the contract .UD1000011D.

REFERENCES