Panoramic Background Model under Free Moving Camera

Naveed I Rao,Huijun Di
Tsinghua University
Pervasive Computing Lab
Beijing, 100084, China,
naveed03,dhj98@mails.tsinghua.edu.cn

GuangYou Xu
Tsinghua University
Pervasive Computing Lab
Beijing, 100084, China,
xgy-dcs@mail.tsinghua.edu.cn

Abstract

segmentation of moving regions in outdoor environment under a moving camera is a fundamental step in many vision systems including automated visual surveillance, human-machine interface, tracking etc. It is also a challenging task due to camera motion, object motion, and outdoor scene challenges i.e. periodic motions of swaying of trees, gradual illumination changes, etc. In this paper, a wide area scene modeling approach for object segmentation under a moving camera is proposed. This approach suffers due to parallax effect, misallignment errors etc and needs their concurrent removal for its success. we explicitly model the dense correspondence between input image and panoramic background model. Foreground segmentation and correspondence estimation are expressed as a unified labeling problem, and are solved efficiently via tree dynamic programming (TDP). Lucas-Kanade method is used to find sparse correspondence between image and model, and robust M-estimator is then applied to find the projective transformation for initialization of TDP's window search. Optimal dense correspondences are achieved and are used to update panoramic background model and as a byproduct, online refined panoramic image is generated which is empty in the beginning and is filled step by step. We test our algorithm with hand-held camera and also with a camera mounted on a moving platform. Experiments proved our algorithm to be robust in performance

1 Introduction

Foreground segmentation is often a fundamental component for many applications such as visual surveillance and human computer interaction. To segment moving object out of image, background subtraction is a typical approach. Especially this approach is helpful when the model is updated over time to cater for the changes in observed information. Normally, the background models are designed for fixed camera and have produced good results [13, 16, 6, 14, 2, 10] . Equally good results are expected incase of moving camera for wide area modeling if achieved successfully. Potential applications may be handheld camera applications, active vision systems e.g. wide area surveillance etc. Since the camera may undergo free motion, it is better to build a panoramic background model to hold pertinent information. To achieve good modeling results, successive images are need to be aligned precisely [1]. However, parallax effect [8], lens distortion, and registration error (caused by noise, inadequate motion model, sub-pixel effect, computational error, etc.), hinders in realizing a stable panoramic background model. It is further complicated in outdoor environment possessing dynamic scenes e.g. abnormal swaying of trees, shimmering of water etc. Without concurrent elimination of these misalignment effects, background model fails due to error accumulation.

There are many works in background modeling, few restricts to fixed camera [16]-[17], others discuss dynamic scene [2, 10, 17, 4, 18] [17] or camera jitter [4][8],[11], but very few deal with moving camera [1, 7, 19]. While extending model to moving camera, the updating stage plays an important role in achieving a stable panoramic background model. However most works do not consider the elimination of misalignment effects during updating (except [19], [11]), thus it is still not a trivial task to prevent error accumulation under panoramic case. Recently, [19] proposes a Spatial Distribution of Gaussians (SDG) model to describe the background under moving camera, and uses the nearest pixel belonging to background inside a neighborhood window to update the model to reduce miss updating effect, but in his model pixels are considered independently. In [11] a joint correspondence based background model is presented which uses estimated correspondence in between current frame and background model for updating, but they only consider small camera jitters under fixed view of field.

In this work, we fully formulate the background modeling problem under panoramic case, and achieve a stable model which can cater for free camera motion, dynamic scenes presents in environment, and above mentioned all misalignment effects within a single mechanism. In order to eliminate error accumulation, dense correspondence between current frame and background model is explicitly modeled and inferred along with foreground segmentation via solving a unified labeling problem. To deal with ambiguities among pixels and obtain a realizable result, piecewise smoothness is enforced on tree-structured spatial relationships of pixels and tree dynamic programming (TDP) is used to solve this labeling problem efficiently. Resultantly, parallax effect and registration errors are reduced significantly, and error accumulation during expanding and updating the model is totally eliminated. As a byproduct, online refined panoramic image is also generated which is empty in the beginning and is filled step by step.

The paper is organized as: Complete framework is explained in section 2. Section 3 describes Jointly foreground segmentation and correspondence estimation. Section 4 explains the updating scheme for the panorama and discussion on experiment results is carried in section 5. Conclusion and future work is presented in last section.

2 Framework overview

In our work, a panoramic background model (PBG) is built over a wide area, with a free moving camera. Due to parallax effect (caused by movement of camera's optical center) and lens distortion, a global transformation between current frame and PBG doesn't exist. Furthermore, because of facts such as noise, sub-pixel effect computational error, and disturbances created by dynamic scenes, the estimation of transformation is also inaccurate. Therefore we only assume an approximate alignment between current frame and model. Our PBG along with associated optimization algorithm compute a dense (pixel level) refined matching between current frame and PBG. This refined matching aid in achieving good foreground segmentation results by eliminating the misalignment error. A specially designed updating scheme of PBG, ensure a stable system over a long period of time. Our framework is shown in figure 1 and the steps are explained below. Step 1: Approximate alignment between current frame and PBG is achieved through projective transformation in initial step. Same as in [1], we always compare current frame with model to estimate the transformation rather than frame by frame comparison to avoid registration error accumulation. A panoramic image of currently explored scene is first generated from PBG by taking the mean value of the most likely model, to enable image based matching. LK [9] is used to find sparse correspondence between current frame and this panoramic image. Since directly searching over whole panoramic image is infeasible for LK (due to time consuming and easily entrapped into local minimum), previously estimated projective transformation is used to give an initialization for LK (formally, use the previously computed transformation to synthesize a virtual image, and then apply LK, as shown in figure 1). Given the estimated sparse correspondence, M-estimator [5][17] is used to calculate the projective transformation between current image and PBG robustly. Step 2: generate an auxiliary image for motion compensation which is used in next steps as an input. Based on the estimated projective transformation in step 1, current frame is transformed into the coordinates of panorama and an auxiliary image is generated by cutting the transformed image out, as shown in figure 1. Further details are given in section 4. Step 3: segment foreground and estimate dense matching from current frame to PBG. Step 4 explains a specially designed stable updating scheme of PBG. Details of above two sections are covered in section 5 and 6 respectively.

3 Joint Foreground Segmentation and Correspondence Estimations

As mentioned in previous section, the correspondences between auxiliary image and PBG are posed as our model parameters. These estimated correspondences provide dense matching, and eliminate the misalignment effects thus improving foreground segmentation. In this section, we will discuss the associated algorithm for foreground segmentation and correspondence estimation. There exists an inherent ambiguity in dense correspondence , a reasonable smoothness constraint is required to enable a matching procedure work accurately. In order to regularize the resulting correspondence, piecewise smoothness assumption of a scene is normally used [12][15], we also apply such a assumption in our problem domain. The rest formulation is similar to [11], except we discuss the panoramic case. For the sake of completeness and with the courtesy of authors, we briefly overview it here, but with our work i.e. PBG point of view. For each pixel in auxiliary image J_t, our goal is to find its correspondence in PBG, as well as classify it as background pixel or foreground pixel. It can be viewed as a labeling problem: assign an optimal label $l = (f, \Delta x, \Delta y)$ to each pixel in J_t while f gives segmentation result (0 for background, 1 for foreground) and displacement vector $(\Delta x, \Delta y)$ award its correspondence. The whole labeling space is

$$L = \{(0, \Delta x, \Delta y)|(\Delta x, \Delta y) \in D\} \bigcup \{(1, *, *)\} \qquad (1)$$

where D is the domain of displacement vector $(x + \Delta x, y + \Delta y)$, $D = \{(i,j)| - a \leq i \leq a, -b \leq j \leq b\}$. Since correspondence is only modeled in-between background pixels, so when a pixel is labeled as foreground (f=1), we do not consider the correspondence for it.

To enforce scene smoothness, a minimal span tree (MST) [3]is generated from an undirected graph which

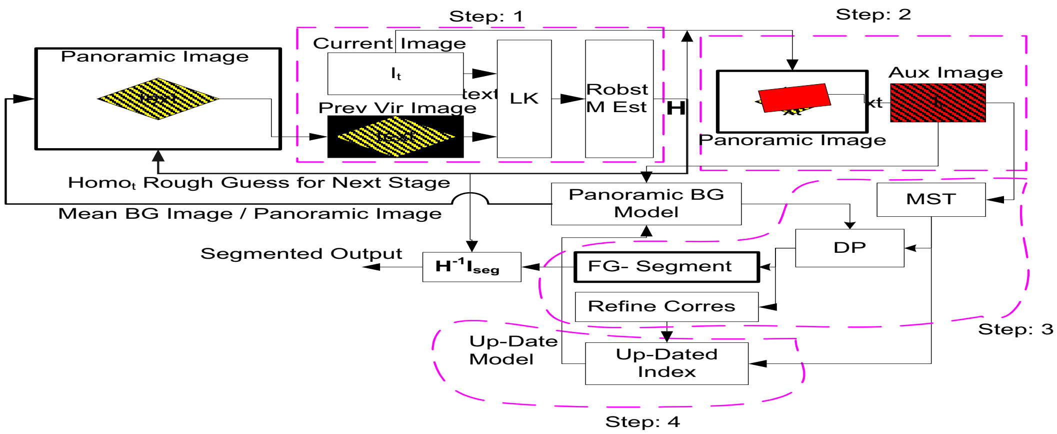

Figure 1. Schematic Diagram of FrameWork

is defined on the auxiliary image J_t: pixels as vertices; edges as piecewise connection between neighbors; and absolute intensity difference between neighboring pixels as a weight of edge to approximate the geometric smoothness. This MST contains most important edges in the graph that smoothness enforcing should be imposed. The Energy function for assigning optimal label is defined as

$$E(l) = \left(\sum_{v \in V} m(l_v) + \sum_{(u,v) \in E} S(l_v, l_u) \right) \tag{2}$$

where V and E are the set of all nodes and edges in MST respectively, l is the conjunction of labels for all nodes, u and v are nodes (also pixels, coordinates in image J_t), l_v and l_u are labels. In eqn. 9,$m(l_v)$ is the data measurement penalizing any disagreement of the label (foreground segmentation along with correspondence) with the observed data (image J_t), and is defined as negative logarithmic likelihood of the label, i.e.

$$m(l_v) = \begin{cases} -\log P_B\left(J(x,y), \Delta x, \Delta y\right), \text{for} f(x,y) = 0 \\ -\log\big((J(x,y), \Delta x, \Delta y)\big) \quad , \text{for} f(x,y) = 1 \end{cases} \tag{3}$$

$S(l_v, l_u)$ in eqn. 4 is the smoothness term defined in the form of Potts model [10]:

$$S(l_v, l_u) = \begin{cases} w_1(I_t(v), I_t(u)), if f_v \neq f_u \\ 0, if f_v = f_u = 1 \\ 0, if (f_v = f_u = 0) \& (\Delta x_v = \Delta x_u) \& \\ (\Delta y_v = \Delta y_u) \\ w_2(I_t(v), I_t(u)), otherwise \end{cases} \tag{4}$$

where w_1 and w_2 are two weights to penalize the discontinuity of labels between parent node and child node in MST.

Due to tree structure, the minimum energy can be written as

$$E_{min} = \min_{l_r \in L} \Big(m(l_v) + \sum_{v \in C_r} E_v(l_r) \Big), \tag{5}$$

where r is the root node of MST, C_r is the set of $r's$ children,$E_v(l_r)$ is defined recursively as

$$E_v(l_p) = \min_{l_v \in L} \Big(m(l_v) + S(l_v, l_p) + \sum_{(u \in C_v)} E_u(l_v) \Big) \tag{6}$$

The minimum energy in equ. 12 can be calculated efficiently via tree dynamic programming [2], and associated global optimal labels are then obtained, accordingly foreground segmentation and dense correspondence are achieved.

4 Updating Panorama

We use the same strategy as [13] to update background model, i.e. the on-line K-means approximation and all pixels are used to update the model, allowing objects to be part of background with the passage of time. As the correspondence is modeled explicitly, the difference is the decision, for which model (which location) should be updated based on the current pixel information. For the pixel (x, y) in input image, suppose the location of the model to be updated is $(x + dx, y + dy)$. After the optimal label $f*$, $\delta x*$, $\delta y*$ obtained for this pixel in section 3, (dx, dy) is calculated as

$$(dx, dy) = \begin{cases} (\Delta x, \Delta y), if f^* = 0 \\ (0, 0), if f^* = 1 \end{cases} \tag{7}$$

Since correspondence is meaningless when the pixel is segmented as foreground, so we update the model at same location (i.e. $dx = dy = 0$). It is worth noting that, since we match input image to the model, different pixels may be matched to the same model. At current frame, background models at few locations may miss their updating, consequentially their evolutions will be slow. Thus foreground object may exist for a longer time than [13] if it is presented during model initialization. However, this does not cause more false segmentation results for the related pixels, as these pixels are matched to other background models and are not detected as ghost.

5 Experimental Results

When a camera is moved with speed, it produces blurring in the consecutive images, due to sensor exposure problem (as shown in fig. 2 (d-e). This blurred images if used for the updating of the model will also blur the whole model. In this part of research, we have used a slowly moving camera to avoid this effect but in future we want to extend this frame work to normal moving camera.

The algorithm is tested on both cases of moving camera around a pivot and slowly free moving. In figure 3, the sequence depicts an outdoor scene of a garden with all artifacts of swaying of trees, illumination changes etc. Second sequence is taken with a handheld camera.

6 Conclusion and Future Work

This work is a part of a project, to perform wide area surveillance using a camera mounted on a pan-tilt platform. In this part ,Background model proposed in [11]-[12] for nominal moving camera is improved and is extended to wide area scenes. Achieved updating results shows blurring at few spots due to n-averaging effect. In its next step , we will reduce this blurring effect and will model foreground pixels explicitly to probe in its effect on the performance.

References

[1] Mittal Annurag and HuttenLocher Dan. Scene modeling for wide area surveillance and image synthesis. In *CVPR*, 2000.

[2] Monnet Antoine, Mittal Anurag, Paragios Nikos, and Ramesh Visvanathan. Background modeling and subtraction of dynamic scenes. In *ICCV*, 2003.

[3] T. Cormen, C. Leiserson, and R. Rivest. *Introduction to Algorithm*. MIT Press, 1990.

[4] A. Elgammal, D. Harwood, and L. Davis. Non-parametric model for background subtraction. In *ECCV*, 2000.

[5] D. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. chap 15, 2003.

[6] I. Haritaoglu, D. Harwood, and L.S. Davis. W4: real-time surveillance of people and their activities. *TPAMI*, 22(8):809–830, 2000.

[7] E. Hayman and J. Eklundh. Statistical background subtraction for a mobile observer. In *ICCV*, 2003.

[8] Kang Jinman, Cohen Isaac, and Yuan Chang. Detection and tracking of moving objects from a moving platform in presence of strong parallax. In *CVPR*, 2005.

[9] B. D. Lucas and T. Kanade. An iterative image registration technique with an application in stereo vision. In *Seventh International Joint Conference on Artificial Intelligence (IJCAI-81)*, 1981.

[10] A. Mittal and N. Paragios. Motion-based background subtraction using adaptive kernel density estimation. In *CVPR*, 2004.

[11] I Rao Naveed, Di Huijun, and Xu Guangyou. Joint correspondence and background modeling based on tree dynamic programming. In *ICPR*, 06.

[12] I Rao Naveed, Di Huijun, and Xu Guangyou. Stereo correspondence using bayesian network and dynamic programming on a color based minimal span tree. In *Advanced Concepts for Intelligent Vision Systems (Acivs 2006)*, 2006.

[13] C. Stauffer and W. Grimson. Learning patterns of activity using real-time tracking. *TPAMI*, 2000.

[14] J. Sun, W. Zhang, X. Tang, and H-Yeung Shum. Background cut. In *ECCV*, 2006.

[15] Olga Veksler. Stereo correspondence by dynamic programming on a tree. In *CVPR*, 2005.

[16] C. Wren, A. Azarbayejani, T. Darrel, and A. Pentland. Pfinder: Real time tracking of the human body. *TPAMI*, 1997.

[17] Quanren Xiong and Christopher Jaynes. Multiresolution background modeling of dynamic scenes using weighted match filters. In *VSNN,IEEE Conference on Multimedia*, 2004.

[18] Sheikh Y. and Shah Mubarik. Bayesian modeling of dynamic scenes. *TPAMI*, Nov.

[19] Ren Ying, CHua Ching-Seng, and Yeong-Kehing Ho. Motion detection with nonstationary background. *Machine Vision and Applications*, 13:332–343, 2003.

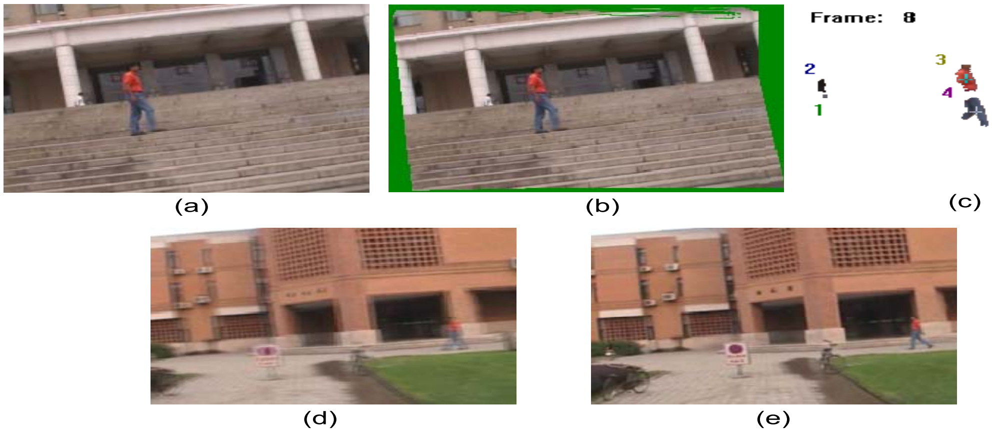

Figure 2. Hand held moving camera:(a) an image from input sequence(b) Mean background image after 8 frames (c) Foreground segmented results: Two consecutive frames taken from a moving camera (d) Camera is moving with speed(e)Camera is moving slowly

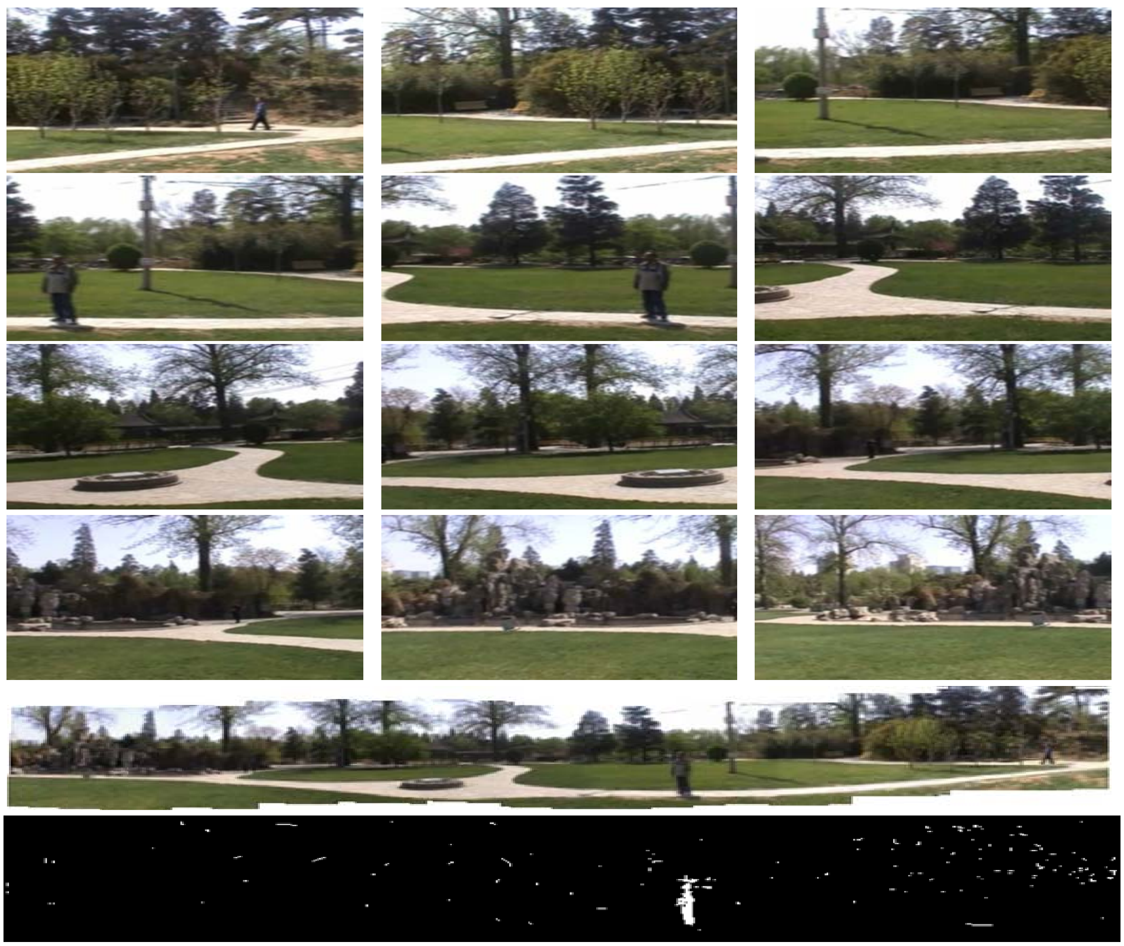

Figure 3. Panorama Generation : Sequence of images from (a) to (k) (starting from top left to right bottom) depicts scene structure from left to right. (l) is Panorama generated based on the images , (m) shows detected motion in it