AN ADAPTIVE BACKGROUND MODEL INITIALIZATION ALGORITHM WITH OBJECTS MOVING AT DIFFERENT DEPTHS

Chia-Chih Chen and J. K. Aggarwal

The University of Texas at Austin
Department of Electrical and Computer Engineering
Austin, TX 78712-0240, USA
{ccchen | aggarwaljk}@mail.utexas.edu

ABSTRACT

Background subtraction is an essential element in most object tracking and video surveillance systems. The success of this low-level processing step is highly dependent on the quality of the background model maintained. Gutchess *et al.* [4] proposed a novel background initialization algorithm that utilizes local optical flow information to locate the stable interval (of intensity values) which is most likely to display background. However, it is found that the accuracy of the computed background is rather sensitive to the parameters used. In addition, their algorithm is not able to handle the scenario where objects are moving at different depths. In this paper, we propose an algorithm which is adaptive to the input sequence and is able to equalize the uneven effect caused by different object depths. Our algorithm is successfully tested on complex indoor and outdoor scenes with promising results.

Index Terms— background initialization, optical flow, multiple foreground depths

1. INTRODUCTION

The separation of moving objects from a static background is a crucial step in applications such as video surveillance, object tracking, and content-based video encoding. The purpose of background initialization is to extract the stationary scene model from a short training sequence where foreground objects may be present. In [3], background is modeled by the median of every pixel's intensity over the sequence. This assumption is true only when each background value appears for more than 50% of the sequence duration. The assumption may be false because the video sequence to be initialized can be selected arbitrarily.

Recently, there have been a variety of approaches proposed for background reconstruction. We may categorize these techniques as unimodal or multi-modal by their model representations. For most unimodal algorithms [1, 3, 4], each background pixel is modeled by a single representative value or distribution. While multi-modal methods [2, 5, 6] maintain several candidate models for each pixel at the same time, which are updated when sufficient evidence of background change is detected. Our background initialization algorithm adopts unimodal representation, which can be updated via batch processing of incoming video frames.

This paper contributes an optical flow based algorithm with the consideration of different object depths and scales of movement. The motion of objects between two consecutive video frames is characterized by optical flow. Gutches *et al*. [4] uses Gaussian-weighted distance to model the influence of local optical flow on the likelihood of background visibility (background likelihood) of each pixel. However, they did not discuss how different scales of foreground activity may affect the selection of parameters. In addition, their algorithm did not consider the effect of multiple foreground depths, which is commonly seen in outdoor scenes. Our algorithm provides a complete solution to these problems.

The algorithm begins with locating time intervals of stable intensities (stable intervals) of each pixel. Optical flow is then computed between every pair of successive frames. The appropriate size of window to search for optical flow depends on the maximum movement among the input sequence, which is approximated under multiple resolutions. The likelihood of a pixel being covered or uncovered is decided by the relative coordinates of optical flow vector vertices in its neighborhood. However, in the presence of multiple foreground depths, different object resolutions can lead to substantial bias in the density of distributed optical flow (Fig. 2(b)). As a result, objects closer to the camera may dominate the value of background likelihood. We equalize the effect of multiple object depths based on local optical flow density. Finally, the value of each background pixel is represented by the median pixel value of the member stable interval whose average background likelihood is the greatest.

2. ALGORITHM

We present the overall description of the algorithm in

Section 2.1. Methods for adaptive parameter estimation and flow density equalization are discussed in 2.2. and 2.3.

2.1. Algorithm Overview

In a video sequence taken from a fixed camera, the background values tend to be constant. To estimate the background successfully, it is assumed that every pixel will display the background for at least a short period of time. For each pixel, stable intervals are defined as non-overlapping intensity subsequences <a_i, …, a_j> that satisfy:

$$\begin{aligned} & j-i \geq d_{min} \\ & |a_r - a_s| \leq \delta_{max}, \forall r,s \end{aligned} \tag{1}$$

In (1), d_{min} (5 frames) is the minimum sequence length required, and δ_{max} (10) is the maximum intensity variation allowed. The search for stable interval is performed and stored for every pixel location.

To estimate the likelihood of background visibility of each pixel, optical flow is computed for every pair of successive frames from the training sequence. However, to capture the motion of objects accurately, an appropriate search range of optical flow is required. The radius of the flow search area around a pixel relates to the magnitude of foreground activity, the estimation of which is discussed in the next subsection. After optical flow computation, assume that n flow vectors are found between a pair of successive frames. These flow vector heads (h) and tails (t) are located at ($x_{i,h}$,$y_{i,h}$) and ($x_{i,t}$,$y_{i,t}$), respectively. For a pixel location (x,y) of frame f-1 and f, the distances from (x,y) to various flow vector vertices in its neighborhood decide the contribution to background likelihood. We model the likelihood contribution of each local flow vertex to a pixel by *weighted* Gaussian-weighted distance. The background likelihood contribution of flow vector heads in the neighborhood N of (x,y) is defined as:

$$\Delta l_f^h(x,y) = -\sum_{i\in N} \frac{w^h(i)}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}[(x-x_{i,h})^2+(y-y_{i,h})^2]} \tag{2}$$

Similarly, the likelihood contribution from local flow vector tails is:

$$\Delta l_f^t(x,y) = \sum_{i\in N} \frac{w^t(i)}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}[(x-x_{i,t})^2+(y-y_{i,t})^2]} \tag{3}$$

The influence of local flow vector heads and tails are weighted by w^h and w^t, the determination of which is detailed in Section 2.3. The contribution of background likelihood between frame f-1 and f is expressed as:

$$\Delta L_f = \Delta l_f^h(x,y) + \Delta l_f^t(x,y), \forall x,y \tag{4}$$

Based on (4), the background likelihood up to frame f is:

$$L_f = \sum_{i=0}^{f} \Delta L_i \tag{5}$$

In the last stage, for every pixel location, the average background likelihood of member stable intervals is evaluated. We define the average background likelihood of a stable interval <f_1,f_2> at (x,y) as:

$$\bar{l}_{<f_1,f_2>}(x,y) = \frac{\sum_{i=f_1}^{f_2} l_i(x,y)}{f_2 - f_1} \tag{6}$$

, where l represents the background likelihood at pixel level. The background value at (x,y) is then displayed by the median pixel value of the member stable interval whose average likelihood $\bar{l}$ is the largest.

2.2. Adaptive Parameter Estimation

The search radius of optical flow, the radius of the neighborhood (N), and the variance of the Gaussian weight function are the three key parameters that affect the quality of the output background model. It is reasonable to assume that the effective values of these parameters are dependent on the scale of foreground movement. For example, in a training sequence with fast moving objects and/or a close camera-to-subject distance, to characterize every object movement, a greater flow search radius is expected. Furthermore, to avoid the loss of background likelihood contribution (Δl) from rapid objects, a larger neighborhood area is necessary. We develop a method to approximate the maximum foreground movement, which is taken as the flow search radius. Our technique is performed under multiple resolutions to ensure accuracy.

To search for the pair of consecutive frames which contains the largest object movement, the more likely pairs are located instead of a specific one. We first calculate the difference image between each pair of successive frames. The difference images are sorted in the descending order by their number of error pixels. The pairs corresponding to the top 10% of the sorted difference images are regarded as candidates. Next, the maximum movement is approximated among candidate pairs. Fig. 1 shows two images of a candidate pair are sub-sampled to build a three-level image pyramid. By assigning an overly large initial search radius, optical flow is calculated between the top level pair (the lowest resolution pair). The length of the longest flow vector is relaxed and scaled up to be the next level search radius.

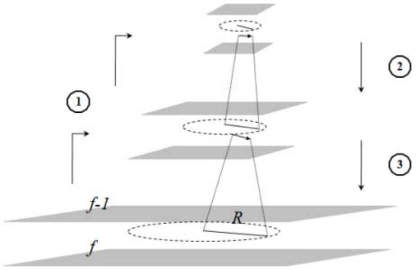
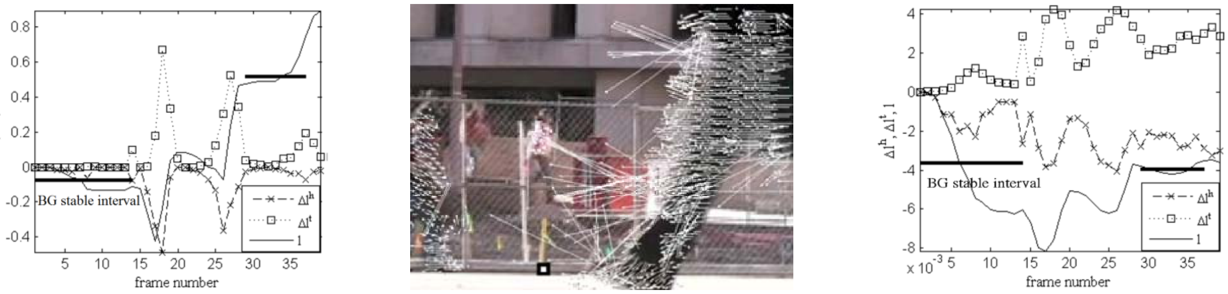


Fig. 1 <f-1, f> is a pair of frames which is likely to contain the largest object movement. This candidate pair is sub-sampled to build a three-level image pyramid (①). The length of the longest optical flow (arrow) from the upper level pair is scaled up to be the search radius (solid line) for the next lower level pair (②, ③). R is one candidate length of the final optical flow search radius.



(a) (b) (c)

Fig. 2 Equalization of the uneven effect of different object depths. (b) shows that different foreground resolutions caused the asymmetric distribution of optical flow. For the marked white pixel in (b), (a) (c) shows the corresponding Δl^h (-x-), Δl^t (-□-), and l (—) as functions of frame number. The two thicker line segments represent the member stable intervals. Without flow density equalization, the approach and departure of the pedestrian closest to the camera correspond to the largest two Δl peaks in (a), which dominate the background likelihood (l). (c) shows the likelihood functions after equalization, the stable interval created by background pixel values (first line segment) gains the greatest $\bar{l}$.

Using the estimated radius, optical flow is computed for the middle level pair. The largest flow length is scaled up to be a candidate of the output radius. Finally, the search radius is determined by the maximum length among all the candidate radii.

From a large number of simulations, we estimated the effective ratios between the parameters as follows:

- The aspect ratio between the radius of neighborhood and the search radius of optical flow is 1.5:1
- The aspect ratio between the radius of neighborhood and the Gaussian standard deviation is 3:1

2.3. Flow Density Equalization

For a video surveillance system, the field of view can be broad and deep, which is especially common for an outdoor setup. Figure 2(b) is an example, due to the wide range of scene depth, the image areas projected by the pedestrians exhibit different resolutions. As a result, the density of the flow vectors discovered from these foreground areas varies. Equation (2) and (3) indicate that the likelihood contribution of local optical flow to a pixel is linear to the number of flow vertices in N. Our purpose is, therefore, to decay the linear relation exponentially so that the effect of uneven local flow density can be attenuated. Let $(x_{i,v}, y_{i,v})$ be the coordinate of a flow vector vertex ($v \in \{h, t\}$). The average likelihood contribution from the same type of flow vertices in its neighborhood (r=1.5σ) can be formulated as:

$$c = \frac{\iint_{x^2+y^2\le r^2} g(x,y)dxdy}{\pi r^2 g(0,0)} \qquad (7)$$

In (7), g is a 2-D Gaussian pdf with mean (0,0) and variance σ^2. Assume that m flow vertices are found in the neighborhood of $(x_{i,v}, y_{i,v})$, the corresponding weight is assigned as:

$$w^v(i) = 1 - \frac{1}{cm+1}\left(e^{\frac{cm+1}{k\pi r^2}} - 1\right) \qquad (8)$$

Fig. 2 shows how flow density equalization corrects the decision on the most likely background stable interval.

3. EXPERIMENTAL RESULTS

Fifteen sequences, indoors (9) and outdoors (6), are used to test the performance of our algorithm. Fig. 3(a) shows the superposition of sample frames from the representative sequences. Table 1 presents the accuracy of our algorithm for the sequences in Fig. 3, which is evaluated by the percentage of correct pixels in the output background. Without any parameter manipulation, the average accuracy of all the testing sequences is *98.69%*. The algorithm achieves slightly lower accuracy in outdoor scenes due to lighting changes and small background motion. We also compare the approximated flow search radius with the maximum foreground movement (measured manually). It is evident that the search radii approximated are adaptive to the training sequences. The estimated search range is long enough to capture any object movement, while not too long to cause erroneous flow search. The average time needed for a 30-frame background initialization is about 25 seconds in MATLAB implementation on a Pentium D PC (2.8GHz).

Table 1. Performance of our algorithm

Image sequence	**Estimated search radius / Maximum movement**	**Accuracy % correct pixels**
Indoor 1	8/6	99.85%
Indoor 2	32/25	99.90%
Outdoor 1	8/5	97.18%
Outdoor 2	36/30	97.00%

4. CONCLUSIONS

In this paper, we present an adaptive algorithm for background initialization. With the help of summarized local optical flow information, pixel-level hypotheses are evaluated to suggest the most likely background interval. To make this idea applicable to wide variety of real-world sequences, we must consider the foreground properties of the input sequence. Depth for example, is one important

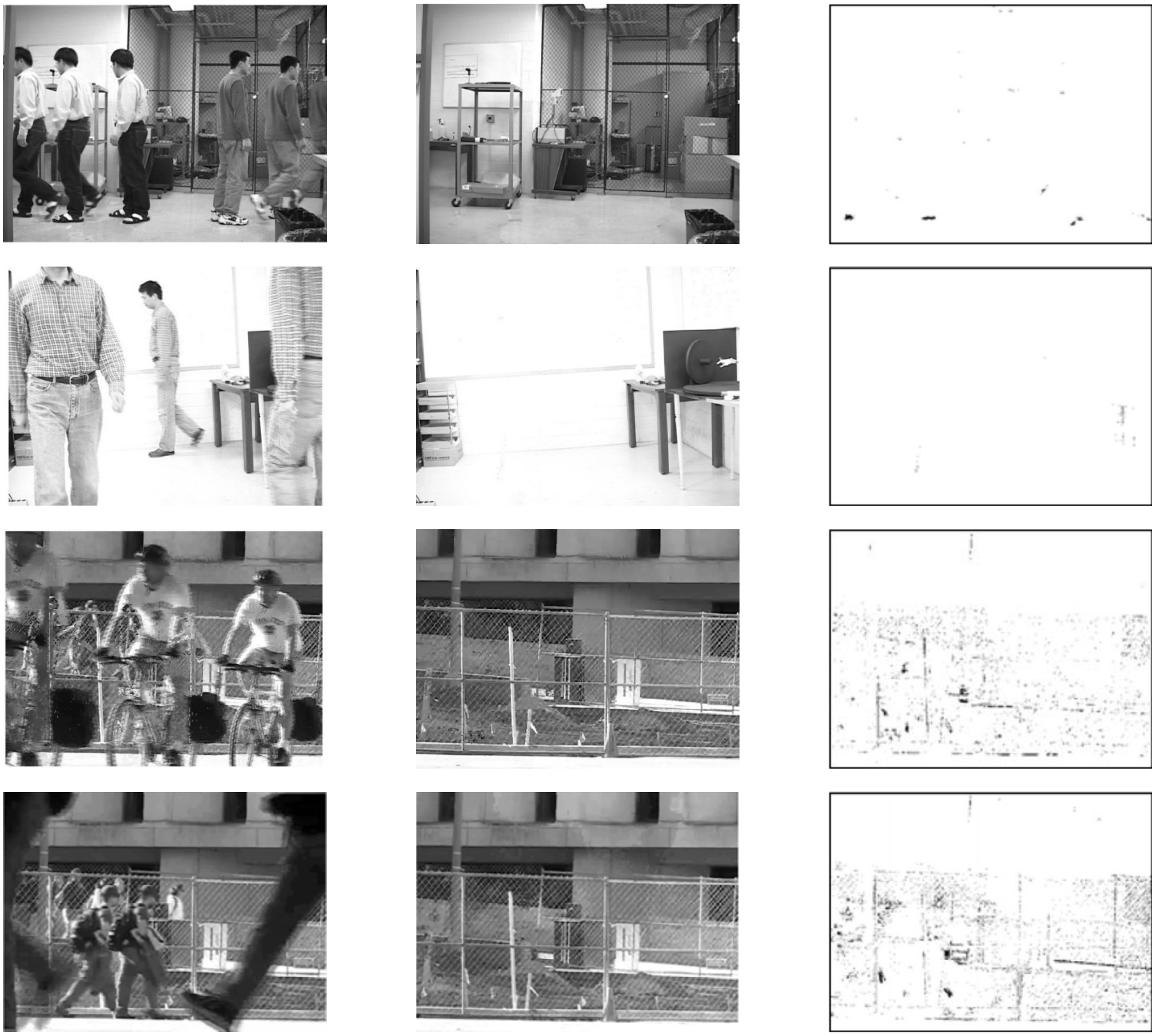

(a) (b) (c)

Fig. 3 Results of the proposed algorithm: (a) snapshots of the representative testing sequences: Indoor 1, Indoor 2, Outdoor 1, and Outdoor 2 (row-wise), (b) the reconstructed background models, (c) maps of error pixels (black, threshold = 25)

factor that decides the likelihood influence of different moving objects. We formulate the desirable weight for each Gaussian weight function so that the influence of different object depths is downplayed. The scale of object movement is another significant factor, which suggests the greatest flow length to be expected. We approximate the largest object movement, and estimate the appropriate values of other parameters based on that. Our method achieves promising results even with complex foreground contents.

5. REFERENCES

[1] K. Dawson-Howe, "Active surveillance using dynamic background subtraction," *Tech. Rep.*, Trinity College, 1996.

[2] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," *Proc. 6th European Conf. on Computer Vision*, (Dublin, Ireland), 2000, pp. 751-761.

[3] B. Gloyer, H. K. Aghajan, K. Y. Siu, and T. Kailath, "Video-based freeway monitoring system using recursive vehicle tracking," *Proc. of Symposium on Electronic Imaging: Image and Video Processing*, 1995, pp. 173-180.

[4] D. Gutchess, M. Trajkovic, E. Cohen-Solal, D. Lyons, A. K. Jain, "A Background Model Initialization for Video Surveillance," *International Conf. on Computer Vision*, (Vancouver, Canada), 2001, pp. 733-740.

[5] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," *IEEE Conf. on Computer Vision and Pattern Recognition*, 1999, pp. 246–252.

[6] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: Principles and practice of background maintenance," *International Conf. on Computer Vision*, (Kerkyra, Greece), 1999, pp. 255-261.