Method of Moving Region Detection for Static Camera

Pravesh Kumar Goel, Paresh P. Kotak, Amit Gupta

B. H. Gardi College of Engineering & Technology, Rajkot, Gujarat, India

goelpravesh2006@yahoo.com

Gujarat Technological University, Principal, AVPTI College, Rajkot, Gujarat, India

kotakp2003@yahoo.com

S. D. College of Engineering & Technology, Muzaffarnagar, Uttar Pradesh, India

toamitji@gmail.com

ABSTRACT

The moving object detection from a stationary video sequence is a primary task in various computer vision applications. In this proposed system; three processing levels are suppose to perform: detects moving objects region from the background image; reduce noise from the pixels of detected region and extract meaningful objects and their features (area of object, center point of area etc.). In this paper; background subtraction techniques is used for segments moving objects from the background image, which is capable for pixel level processing. Morphology operation (Erosion and dilation) are used to remove pixel to pixel noise. In last level, CCL algorithm is used for sorts out foregrounds pixels are grouped into meaningful connected regions and their features.

Indexing terms/Keywords

Background modeling, background subtraction, Pixel level processing, foreground detection.

Academic Discipline And Sub-Disciplines

Computer Science and Image Processing.

SUBJECT CLASSIFICATION

Image Processing.

TYPE (METHOD/APPROACH)

Provide examples of relevant research types, methods/algorithm and approaches for this field: E.g., Literary Survey, Experimental Studies, System Model etc..

Council for Innovative Research

INTRODUCTION

Moving object detection is a difficult task in a complex environment because of moving object interact with each other and it may move in unexpected ways. This requires a robust method without being affected by changes of environment features. Adaptive background subtraction method [1] is assumed to work real time as a part of a video-based surveillance system. The computational complexity and even the constant factors of the algorithms are important for real time performance. The system is initialized by feeding video imagery captured by a static camera. To detect any moving object from the incoming frame; the system is divided into three main stages: detection; reduce noise and extract meaningful objects and their features (area, width, height and center of a region). Detection stage is able to differentiate moving foreground objects from static background objects in dynamic scenes. Morphology operation is used to remove noisy pixels from the foreground pixels map that are not a part of foreground regions and to remove the noisy pixels from the background pixels map that are close to and inside object regions that are actually foreground pixels.

Most of the methods are used for both color and gray scale video imagery. The first step is distinguishing foreground objects from stationary background. To achieve this, uses the adaptive background subtraction and low-level image processing methods to create a foreground pixel map at every frame and then extract the features such as bounding box, area and center of mass of individual object from group of connected regions in the foreground pixel map.

RELATED WORK

There are several methods for moving object detection like background subtraction is a commonly used technique for motion detection in static scenes [2]. In [3] Heikkila and Silven uses subtracting the current image pixel-by-pixel from a reference background image for detect moving regions. In Lipton et al [4] uses a temporal differencing method takes consecutive video frames and determines the absolute difference. Collins developed a hybrid method that combines three-frame differencing with an adaptive background subtraction model [1]. The Haritaoglu et al:W4 [5] system uses a statistical background model where each pixel is represented with its minimum (M) and maximum (N) intensity values and maximum intensity difference (D) between any consecutive frames observed during initial training period where the scene contains no moving objects. Wren, et al. [6], each point in a scene estimated mean intensity value by using a Gaussian distribution. Most of the optical flow methods are computationally complex and very sensitive to noise and cannot be used real-time without specialized hardware [7]. As noted by Toyama et al. [8], Elgammal et al. [9] and Harville et al. [10], there are several problems that must be correctly detect moving objects by a good background removal algorithm. A good quality background subtraction algorithm should be able to handle non-stationary background objects e.g. waving trees, and image changes due to camera motion. A background removal system should adapt to illumination changes whether or sudden changes, whether global or local changes such as shadows and inter reflections.

SYSTEM OVERVIEW

In visual surveillance systems; the first step is detecting foreground objects from video sequence. This is useful for higher processing levels such as tracking; classification and behavior analysis for reduce processing time. When the dynamic scene changes such as light reflectance, repetitive motions for example waiving tree leaves, camera noise, shadows and sudden illumination; it makes a reliable system but it difficult for fast object detection. The system diagram of moving object detection method is shown in figure 1. Object detection method depends on a six stage process to extract objects with their features from video sequence. The first step is the background scene initialization. There are various methods used to background model of the scene. Next step is detecting the foreground pixels map by using the process of background subtraction model from current image from video sequence. To achieve this, use a combination of background subtraction and low-level processing methods to create a foreground pixel map at every frame. The pixel-level process is dependent on the background model that is initialized in starting of this purposed system and when dynamic scene changes the background model is used to update. Also, due to camera noise or environmental effects the detected foreground pixel map contains noise. Pixel-level post-processing operations are performed to remove noise in the foreground pixels.
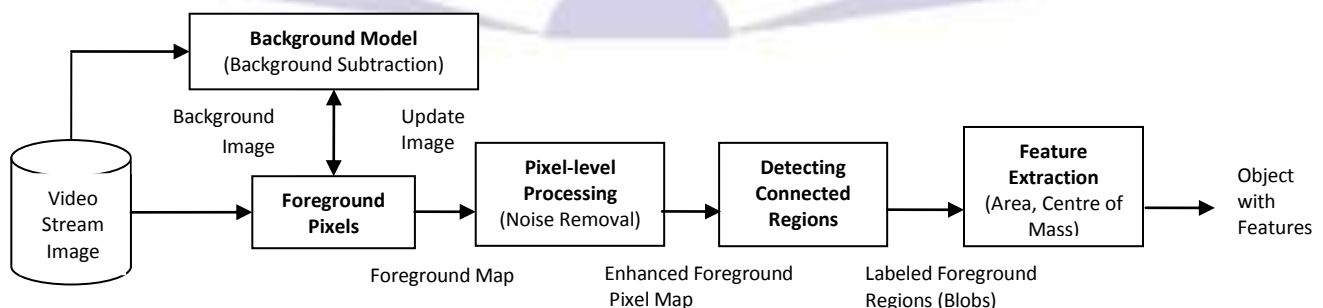
Fig 1: System block diagram of Moving Object Detection

Once get the filtered foreground pixels, in the next step, connected regions are found by using a connected component labeling algorithm [11] and objects bounding boxes are calculated. The connected regions are close to but disjoint due to defects in foreground detection process. Some relatively small regions caused by environmental noise are eliminated in the region-level post-processing step. In the final step of the detection process, a number of object features are extracted from current image by using the foreground pixel map. These features are the area, centre of mass etc. of the regions corresponding to objects that may be use in further processing.

MOVING OBJECT DETECTION

Input Video

This system using input as a video as shown in the system block diagram. A video stream is recorded by digital camera. A video convertor software is used for convert a video into uncompress AVI format file.

- Video: Codec- Cvid, 320 x 240 pixel resolution and 25 frames/second.

- Audio: Codec-WMA2, Bit Rate-320008, Sampling Rate-44100/2.

The sequence of the frame retrieves from the uncompress AVI format video and then the captured frames are converted into grayscale image, which are inputs for the system block diagram.

Background Model

According to Cristani et al. [12], a background modeling is a process that is dividing in three model; representation, initialization, and adaptation model. The first model used for represent the background; the second one used for initialization of this model, and the third used for adjusting the model to the background changes (e.g. suddenly illumination changes). Generally background models have two distinct stages in their process: initialization and update.

Foreground Detection

To create a foreground pixel map; foreground detection methods compares current frame sequence with the background model and then apply low-level image post-processing methods to remove camera noise or environmental effects the detected foreground pixel map contains noise and extract object features at every video frame. Foreground pixels map obtained usually in binary form is "0' represent by background pixel map and "1" represent by foreground pixel map. For detecting the foreground region this system uses Adaptive Background Subtraction Method. In the following subsections we will discuss first how to generate the difference pixel map given the background model and the current frame, and then we will discuss the correct threshold value. The correct threshold value depends on the camera noise, on the scene, and on the illumination conditions.

Adaptive Background Subtraction Model: Background subtraction algorithm [1, 16] is commonly used technique for motion segmentation. In this method, at the start of the system a reference background is initialized with the first frames of video. For every new frame, the foreground pixels map are identified by subtracting the intensity values of reference background from intensity values current frame. The foreground pixels maps are stored in new frame and compare with a predefined threshold value pixel by pixel. When dynamic scene changes the reference background and the threshold values are updated by using the foreground pixel information.

Let $G_i(x, y)$ represents the gray-level intensity value at pixel position (x, y) of video image sequence G which is in the range [0, 255]. Let $BG_i(x, y)$ represents background intensity value for pixel position (x, y) which is estimated over time from video images $G0$ through G_{i-1}. The current video image at pixel position (x, y) belongs to foreground image if it satisfies given equation 1:

$$|G_i(x, y) - BG_i(x, y)| > Th_i(x, y) \qquad (1)$$

Where $Th_i(x)$ is a threshold value estimated using the image sequence $G0$ through G_{i-1}. The value for threshold [13] becomes very important because:

- If the threshold is too low, background brightness could be sudden increase that could be false detection.

- If the threshold is too high, a moving object with brightness close to the background will not be detected as shown in figure 2.

Equation (1) is used to generate the foreground pixel map which represents as a binary array where "1" corresponds to a foreground pixel and "0" stands for a background pixel. It takes very less memory for performing the calculation. The reference background $BG_i(x,y)$ is initialized with the first video image $G0$, $BG0 = G0$ and the threshold image is initialized with some pre-determined value . In this system we consider pre-determined value of threshold is 35.
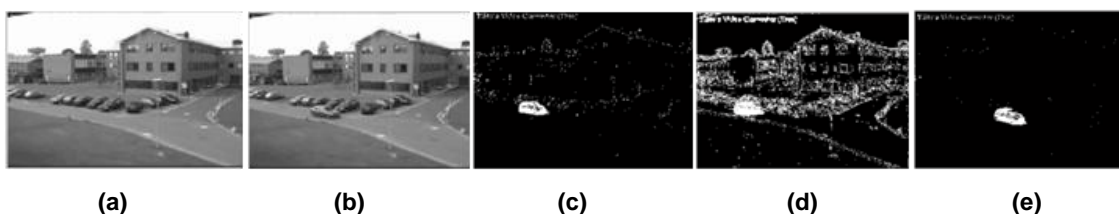

Fig 2: (a) background Image (b) Current Image (c) Detected region (d) Detected region with low Threshold (e) Detected region with high Threshold

This system can be used in outdoor environments and indoor environments. Therefore reference background images and threshold images needs to update itself with incoming images to the dynamic changes such as global illumination change. The update process is different for pixel positions which are detected as a foreground pixels or as a background pixels if it is satisfies as per given equations 2 and 3:

$$BG_{i+1}(x,y) = \begin{cases} \alpha BG_i(x,y) + (1-\alpha)G_i(x,y), & x,y \in \text{non moving region} \\ \beta BG_i(x,y) + (1-\beta)G_i(x,y), & x,y \in \text{ moving region} \end{cases} \quad (2)$$

$$Th_{i+1}(x,y) = \begin{cases} \alpha Th_i(x,y) + (1-\alpha)(\gamma * |G_i(x,y) - BG_i(x,y)|), & x,y \in \text{non moving region} \\ Th_i(x,y) & x,y \in \text{ moving region} \end{cases} \quad (3)$$

Where α, β and γ are a constant value that's belongs to |0.0,1.0| which give us how much information from the incoming image are place to the background and threshold images. In other words, if each background pixel is considered as a time series, the background image is a weighted local temporal average of the incoming image sequence and the threshold image is a weighted local temporal average of γ times the difference of incoming images and the background. The values for α, β and γ are experimentally determined by examining several indoor and outdoor video clips. The value of α cannot be consider too small because it may create an artificial "tails" behind of moving objects. So to prevent artificial "tails" forming behind moving objects need to kept large value of α as shown in figure 3.

(a)   **(b)**
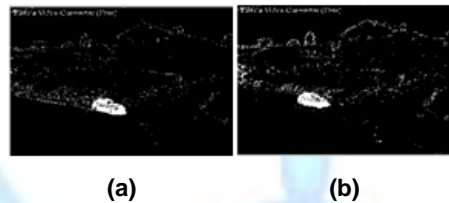
Fig 3: (a) Detected region with α = 0.2 very small (b) Detected region with α = 0.9 very high

If β is too small, foreground objects will be merged to the reference background and it will give inaccurate segmentation in later frames. Objet detecting process will be stop because of moving objects will not be possible to detect. If β is too big, objects may never be merged into the background image, thus background model would not adapt to long-term scene changes. In equation 2 where β=1.0 is corresponding to the background update process presented in [14].

PIXEL LEVEL POST-PROCESSING FILTER

The outputs of background subtraction algorithms generally contain noise due to camera noise, reflectance noise, shadows and sudden illumination change and background colored object noise cannot be handling by background model and affects the outputs at many calculation stages during the processing of a frame and becomes inaccurate frame due to noise. Few algorithms are implemented to improve the image quality to detect the moving object.

Median Filter

The median filter is normally used to reduce noise in an image and it is a simple and very effective noise removal filtering process. Noise is removed by replaces the center value with the median of all the pixel values in the matrix. If we considering an example of 3x3 matrix [15].

Table 1. (a) original image (unfiltered image). (b) median filter (after replacing center value)

6	2	8
3	9	4
1	5	7

*	*	*
*	5	*
*	*	*

The median filter sorts the value of the given matrix and then median value is replace by center value. The sorted values are 1, 2, 3, 4, 5, 6, 7, 8, 9 and median value 5 will be replaced by centre value in the matrix. This process is performed over the whole image and reduces the noise.

Morphological Operations

We apply Morphological Operations, erosion and dilation [16] to eliminate noise form foreground pixels map and also used to eliminate the noisy background pixels inside or near object regions that are mainly foreground pixels. With the help of structure element and set operator such as intersection, union etc. morphological operations work usually on binary images. Generally structuring element finds the details of input images operations to be performed. The size of structuring element is 33 and the centre pixel represents its origin. It is applied over the image and at each pixel of the image and structuring elements is compared with the ones on the image. There are mainly two basis operations of morphology approach, first, translation of structuring element over the image, second, are the erosion and/or dilation of image content. With the help of Morphological Operations, the shape of an image can by analyze and manipulates by marking the area where the structuring element fits. There are many morphological operations but here we are used two basic operations; erosion and dilation operations.

Erosion: It erodes thick boundary pixel of foreground regions. (i.e. white pixels). Therefore areas of foreground regions shrink in size, and holes with those areas become larger.

Dilation: It expands the foreground region boundaries with one pixel. The basic effect of the operator on a binary image is to enlarge the boundaries of foreground regions (i.e. white pixels). Thus areas of foreground pixels grow in size while holes within those regions become smaller.

The erosion and dilation operator takes two pieces of data as inputs. The first is image which is to be eroded or dilated and the second is set of coordinate points known as a structuring element. The morphological filters applied in order and amounts of these operations to get suitable image. The order of these operations affects the quality of image and the amount affects both the quality and complexity of noise removal. For example, if we apply dilation followed by erosion we would dilate one- pixel thick isolated noise regions; this order will successfully eliminate some of the non-background noise inside object regions. In case we apply these operations in reverse order, which is erosion followed by dilation, we would eliminate one-pixel thick isolated noise regions but this time we would not be able to close holes inside object. To solve this problem, we move one step further by filling these holes using the morphological operation "fill". It fills isolated interior pixels (individual 0's that are surrounded by 1's), such as the center pixel in given pattern:

1	1	1
1	0	1
1	1	1

Dilation process has many good criteria such as it can repair the broken edges, help in getting smoother border etc, but its drawback is when applying on a small object. The following steps below have been applied in order to obtain better results.

- Calculating the entire area of the moving object.

- If area of the moving region in the foreground pixels map is greater than or equal to 500 then dilation process will be applied over moving region otherwise dilation process will not be performed. Figure 4 shows the result of pixel-level post-processing of the initial foreground pixel map.

| (a) | (b) | (c) | (d) |

Fig 4: Pixel level noise removal sample. (a) Background Image (b) Current Image (c) Detected foreground regions before noise removal (d) Foreground regions after noise removal.

DETECTING CONNECTED REGIONS ANALYSIS

After detecting foreground regions and applying pixel level post-processing operations to remove noise, the filtered foreground pixels are grouped into connected regions and labeled. By using a two-level connected component labeling (CCL) algorithm [16, 17] extract the area of each object, number of moving objects in the scene and the bounding boxes of these objects (width, height and center). After finding individual blobs with the help of area, width, height and center of the detected regions that correspond to objects, the bounding boxes of these regions are calculated as shown in figure 5.

| (a) | (b) | (c) |

Fig 5: (a) background Image (b) Current Image (c) Current Image with Bounding Box.

PROPOSED ALGORITHM

1. Read color Audio/Video uncompressed AVI file.

2. Initializes a Reference Background by first frame of the Input Video.

3. Convert RGB frame/ Reference background frame/ color map into grayscale frame.

4. Initialize Threshold value with some pre-determined value.

5. FOR (current frame :: last pixel position)

   a. Convert current frame into grayscale frame.

   b. FOR (current pixel position=initial pixel position: last pixel position)

      Calculate difference of pixels values of background frame and current frame.

Store the absolute value of difference of two frames in new frame at same pixel position.

END FOR

c. [Frame Difference is filtered with predefine threshold value]

FOR (current pixel position=initial pixel position: last pixel position)

IF frame difference > threshold value

foreground map =1

ELSE

foreground map=0

END IF

END FOR

d. [Update the Background image and Threshold Values]

FOR (current pixel position=initial pixel position: last pixel position)

IF(foreground map ==0)

Background image= α* Background image +(1- α)* foreground map

Threshold image= α*thresh value +(1- α)*(γ * new frame)

ELSE

Background image = Background image;

Threshold image = thresh value;

END IF

END FOR

e. Apply median filtering to reduce "salt and pepper" noise from foreground pixel map.

f. Remove small area Object from foreground pixel map.

IF areaObj > 500

Create morphological structuring element (STREL).

Clean up Frame using Morphology operation.

g. Calculate the Center point, Height and Width of the Mass.

h. Calculate the bounding box over detected mass.

i. Store the detected regions in array for further processing.

END FOR

6. Display the detected object.

EXPERIMENTAL RESULTS

To test above algorithm, on an up to date Intel Core(TM) 2 Duo CPU at 2.20 GHz with 1 GB of RAM, running under windows XP. The video sequence is obtained from static camera with different views. The technique has been tested on more than 5 video shots containing a total of 3071 images including sequences with noise. Every video has single human, group of human and vehicles of different types. Out of 5 video shots it failed to detect moving objects accurately for one video because of the small size of detected object in the video and rest of the detection results were accurate without any change in parameters. For purposed system we used indoor and outdoor real environments videos. Some results from different videos are shown in figure 6.

Fig 6: Filtered foreground pixels and connected regions with bounding boxes

CONCLUSION AND FUTURE WORK

In this paper; Background subtraction method is implemented for detecting moving region from the stationary videos. The Background subtraction method is capable to give satisfactory results in terms of detection quality but there is no perfect algorithm to detect moving object. A perfect algorithm system should be capable to solve many problems such as moved objects segmentation, shadows removal, suddenly change of illumination, tree waving and so on. But in this work we try to enhance it result and reduce inaccurate segmentation complexity using median filters and morphological operations to remove paper salt noise, camera noise and environmental effects noise. Detected regions would be used for higher level pixel to pixel processing.

ACKNOWLEDGMENTS

REFERENCES

[1]   R. T. Collins et al. A system for video surveillance and monitoring: VSAM final report. Technical report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May 2000.

[2]   A. M. McIvor, "Background subtraction techniques," In Proc. of Image and Vision Computing, Auckland, New Zealand, 2000.

[3]   J. Heikkila and O. Silven, "A real-time system for monitoring of cyclists and pedestrians," In Proc. of Second IEEE Workshop on Visual Surveillance, pp. 74– 81, Fort Collins, Colorado, June 1999.

[4] A. J. Lipton, H. Fujiyoshi, and R. Patil, "Moving target classification and tracking from real-time video," In Proc. of Workshop Applications of Computer Vision, pp. 129–136, 1998.

[5] I. Haritaoglu, D. Harwood, and L. Davis, "W4: A real time system for detecting and tracking people," In Computer Vision and Pattern Recognition, pp. 962–967, 1998.

[6] C. R. Wren, A. Azarbayejani, T. J. Darrell, and A. P. Pentland, "Pfinder: Realtime tracking of the human body," IEEE Pattern Recognition and Machine Intelligence, 19(7):780-785, July 1997.

[7] L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis," Pattern Recognition, 36(3):585-601, March 2003.

[8] Kentaro T, John K, Barry B, Brian M. Wallflower: Principles and Practice of Background Maintenance, ICCV, Seventh Int. Conf. on Com Vision (ICCV'99) - 1999; 1: 255-261.

[9] Elgammal A, Duraiswami R, Harwood D, Davis LS. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. Proc. IEEE Jul. 2002; 90(7): 1151- 1163.

[10] Harville M., Gordon G., Woodfill J. Foreground Segmentation Using Adaptive Mixture Models in Color and Depth. event, IEEE Workshop on Detection and Recog of Events in Video (EVENT'01), 2001; 3-12.

[11] F. Heijden. Image Based Measurement Systems: Object Recognition and Parameter Estimation. Wiley, January 1996.

[12] Cristani M., Bicego M., Murino V. Multi-level background initialization using Hidden Markov Models. In First ACM SIGMM Int. workshop on Video surveillance 2003; 11-20.

[13] Inigo, R.M., "Application of machine vision to traffic monitoring and control", IEEE Transactions on Vehicular Technology, 38 (3) (1989) 112-122.].

[14] S. Khan, O. Javed, and M. Shah, "Human tracking in multiple cameras," in 8th IEEE International Conference on Computer Vision, ICCV 2001, Vancouver, Canada, July 2001.

[15] Wang, L. Hu, W. and Tan, T. (2003). Recent developments in human motion analysis. Pattern Recognition, 36(3):585–601.

[16] Yiğithan Dedeoğlu, B. Uğur Töreyin, Uğur Güdükbay, and A. Enis Çetin, "Silhouette-Based Method for Object Classification and Human Action Recognition in Video", T.S. Huang et al. (Eds.): HCI/ECCV 2006, LNCS 3979, pp. 64–77, 2006, Springer-Verlag Berlin Heidelberg 2006 .

[17] R. C. Gonzalez and R. E. Woods. Digital Image Processing. Pearson Education, 2002.

[18] F. Heijden. Image Based Measurement Systems: Object Recognition and Parameter Estimation. Wiley, January 1996.

[19] http://www.mathworks.com/applications/imageprocessing/

[20] http://www.andrewsenior.com/technical/surveillanceclass/

Author' Biography

Praveshkumar Goel is working as an Assistant Professor in Information Technology Department at B. H. Gardi College of Engineering & Technology, Rajkot, Gujarat, India. He is having Life Time Membership of the Indian Society for Technical Education (ISTE) and active Nominee Member of CSI Branch (Computer Society of India). He is having more than 7.0 years of experience in Academic and Company. He published papers in INTERNATIONAL JOURNALS/ INTERNATIONAL CONFERENCES/ NATIONAL CONFERENCES. He attended NATIONLACONFERENCE/ WORKSHOP in different areas. He received his M. Tech. in Computer Science & Engineering from Nirma University, Ahmedabad, Gujarat. His research interests are in Digital Image Processing and Computer Networks.

Dr. Paresh Kotak is working as a principal, AVPTI College, Gujarat, India. He received his Ph.D. in computer engineering from Saurashtra University, Rajkot, Gujarat, India. His research interests are in Intelligent Computer Networks and Image Processing. He is having more than 20 years experience and publish numbers of paper in INTERNATIONAL JOURNALS/ INTERNATIONAL CONFERENCES/ NATIONAL CONFERENCES.

Prof. Amit Gupta is an associate professor in Computer Science & Engineering department at S. D. College of Engineering & Technology, Uttar Pradesh, India. He received his M. Tech. in Computer Science & Engineering from GGSIP University, Delhi, India. His research interests are in Image Processing, Software Engineering, and Information Security. He is having more than 12 years experience in academic.