TRACKING AND DETECTION OF OCCLUDED HUMAN

Thungamani A N[1], Mr. Rajashekar K J[2]

[1]CSE, Kalpataru Institute of Technology, (India)

[2]ISE, Kalpataru Institute of Technology,(India)

ABSTRACT

Tracking of human is a common requirement for many real world applications such as video surveillance, games, cultural and medical applications. In this paper an appearance based tracking system is proposed which tracks human motion from different video scenes. The system is based on two main models such as color histogram and color correlogram model. This system works well in indoor and semi-indoor environment. The proposed system comprises of five major steps. Firstly all motion blobs in the video scene are extracted. Morphological operations are applied then on extracted motion blobs for removing noises associated with blobs. After that all the motion blobs are counted and labeled by connected component finding. Correlogram and histogram model for each blob are built from the color information of moving blobs in the next step. Similarity measures between the blobs of the current image frame and previous image frame are calculated. Finally system performs tracking based on the similarity measures from frame to frame. Multiple people can be tracked consistently by this proposed system. The system is consistent during partial occlusion. During occlusion occluding entities are tracked as one entity and tracked separately when they split.

Keywords: Computer Vision, Video Surveillance, Background Subtraction, Color Histogram, Color Correlogram.

I. INTRODUCTION

The analysis of human actions by computer is gaining more and more interest. Any system that needs to employ automated surveillance on a scene must require analyze the happenings on the scene firstly. After analyzing the scene properly system can be able to take any necessary decision based on those gained information. So in any sort of application related to automated surveillance or analyzing scene must go through a precise pathway of motion tracking of objects on the scene. There is particular area in human brain that processes what we accept as image in the form of electromagnetic wave. As a basic idea of theological speculation, scientists have always been trying to build something which would act something similar like human being. The ultimate goal is to make a machine capable of doing something similar enough to how human being deals with surroundings. According to a survey carried by Oklahoma State University about 83% of information about surroundings we do intake through our sight [12]. That's where the computer vision comes into actions, which deals with the mechanisms how a machine can be made with sight sense. With the increase threats of terrorism video surveillance system on important public places may greatly mitigate the losses embodied by terrorism. With the competitive environment of business and technology at present time surveillance system is needed in private security level to provide more secure and protected environment. As a whole human motion tracking firstly deals with detection of moving objects [1]. There are extensive approaches on human motion tracking [1, 2]. Presented approach is developed based on color information that is appearance of human motion entities. This system is capable of tracking each moving human on the video scene. Besides system has a total control on each moving blob on the video so that this control could be used in any kind of higher level analysis (activity recognition, human machine interaction). The system comprises of background subtraction, connected component finding, building models for moving entities and finally tracking the peoples based on built models. Proposed system is robust to gradual illumination changes because of application of adaptive update to the background and the color models. The system is developed assuming that the only moving entities on the video scene is human. The work is based on focusing indoor or semi -indoor environment, with static camera. But the system could easily be incorporated into outdoor environment.

II. RELATED RESEARCH

There are two vital phases of human motion tracking system. First one is the foreground segmentation. The most vital part of our proposed surveillance system is the extraction of foreground objects. Two mainly common techniques exist in the literature are the optical flow computation and background subtraction. Background subtraction works with calculating the differences between the foreground pixel and corresponding background pixel and with a help of threshold value final decision is taken. There are number of systems where background subtraction was adopted for foreground segmentation [1, 2, 4]. In [1] background subtraction was done by modeling, representing and maintaining background which is time consuming for special cases such as indoor environment with symmetric lighting all over the scene. Maintenance of statistical background used in [2], may flaw in gradual illumination with changing environment. Optical flow based methods have been used in [5]. As our main focus on to the tracking of human motion, we proposed a simple equation for background subtraction. Addressing tracking position is also an important portion of human man motion tracking. Several work have done for tracking purpose. Mean shift based methods used in [11] may flaw for lack of proper kernel choice or proper target model build up. Here in this paper, color based histogram and correlogram models are used for tracking. This method is very robust in case of rotation or challenging gradual illumination changes.

III. ARCHITECTURE

A total glance of the proposed system is given in the Fig 1. Brief description of each of the cited phases in the figure is discussed in the following sections with result.

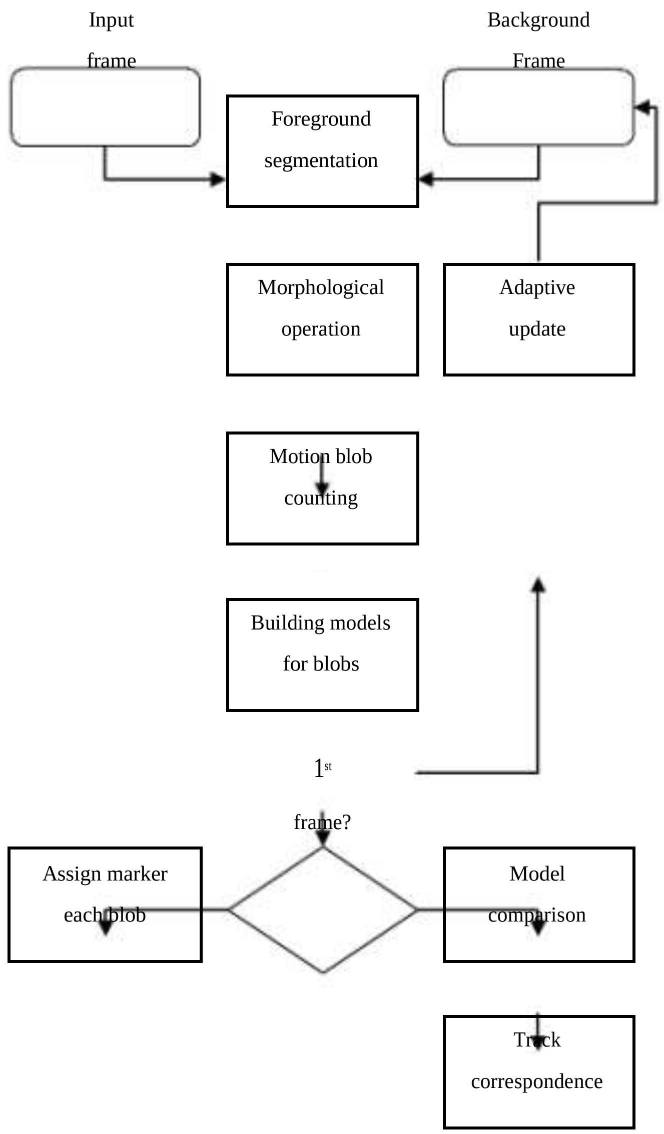

Fig. 1. System architecture of Human Tracking System

3.1 Foreground Segmentation

Background subtraction method is used here to extract foreground objects from current image frame. From the input background image frame and current image frame the following equation is used to extract foreground objects,

$$FI_{(x,y)} = \quad [(r_f - r_b)^2 + (g_f - g_b)^2 + (b_f - b_b)^2] \qquad (1)$$

Where $FI_{(x,y)}$ is the extracted foreground pixel and r_f, g_f, b_f are the (x,y) pixel's RGB value of current image frame and r_b,

g_b, b_b are the corresponding pixel's RGB vector of background image. Then a threshold value is used to find out the correct foreground motion entities. Threshold value for the background subtraction is chosen statistically from a set of tested historical thresholds. The system is run for a large number of test cases using heuristic threshold value for each test case. Finally from those values of heuristic threshold, mean value of threshold set is chosen which gives minimum standard deviation. This optimal threshold value is then used in the system.

3.2 . Morphological Operations

Some unexpected noises appear due to unusual camera movements, gradual illumination changes. Holes are created inside extracted moving blobs which must be constructed to make the system's performance accurate. Image closing operation is applied to extracted foreground entities. Image closing operation is defined by an image dilation operation followed by an image erosion operation with same structuring element. As a structuring element a line vector of 10 pixels is used with 80 degree angles in the developed system. Elliptical or polygon shaped structuring element might grow the motion entity unusually at the edge. That's why vector line structuring element was chosen for this morphology.

3.3 Motion Blobs Counting

Counting how many objects in the scene is the most crucial part of the system. All the connected components must be detected and controlled successfully for better tracking. Breadth First Search (BFS) is used to find all the connected components in the frame. Unexpected noises or existed blobs having no possibility of being human entity are removed in this phase based on a size thresholding. By the end of this phase system is able to answer the questions like how many people on the scene, how many pixels each human object consists of etc. The system also labeled each human for the purpose of tracking in future in this phase.

3.4 Appearance Based Model

System is based on appearance based color histogram and color correlogram models. Models for each detected human entity are built dynamically.

1) Color histogram: Color histogram is the frequency of color in an image showed in a structured way. Formally The color histogram h of image I is defined for i ∈ [m] such that $hist_I(c_i)$ gives for any pixel in I , the probability that the color of the pixel is c_i could be defined as,

$$hist_I(c_i) \quad \frac{count_I(c_i)}{|size_I|} \qquad (2)$$

Where, $count_I(c_i)$ is the count of pixels of color c_i,

$size_I$ is the size of the image I. Color histogram model are built for each detected moving blob in the video scene and by

comparing models of current image frame with models of previous image frame tracking is achieved.

2) Color correlogram: Unlike histrogram color correlogram has some corretional information about frequency of color. Informally correlogram is the probability of a pixel to be some color from a pixel of some color in a given distance.

Formally Correlogram value of a pixel of color c_i distance k to be of color c_j can be expressed as,

$$correl_I(c_i, c_j, k) = \frac{count_I(c_i, c_j, k)}{8k * count_I(c_i)} \quad (3)$$

Where $count_I(c_i, c_j, k)$ symbolizes the count of total of a pixel of color *ci* is *k* unit distant from another pixel of color *cj*. And 8k is a factor representing how many neighboring pixels in distance k. if *k*=1 then, 8 neighboring pixels, if k=2 then, 16 neighboring pixels and so on.( ) *I i count c* is the count of pixels of color ci in the image I. Correlogram models are also used for each moving blobs as histogram models. Main goal of using correlogram having still histogram is make the system robust to rotational changes of moving blobs from frame to frame.

3.5 Similarity Measure

Similarity measure is required among the models of current image frame's moving blobs and previous image frame's models for implementing the tracking from frame to frame. Once models have been acquired a good similarity measure is to be applied to get correct result. In this system a modified version of L_1 distance norm is used. Suppose that we have *HP* and *HC*, the color histograms of previous and current frame respectively. Then our similarity result can be represented as,

$$SIM = \frac{\sum_{i=1}^{256}[(HP_i - HC_i)]}{\min(NC, NP)} \quad (4)$$

Where *NP* is the number of pixels in the blob of previous frame, *NC* is the number of pixels in the blob of current frame.

The similarity measure of correlogram could be similarly represented.

3.6 Tracking

We have a value of each of the blobs of current image frame with each of blobs of previous image frame from equation (5). Based on this value and with comparing it with some threshold tracking decision are made. Threshold value may differ from the fact which one of histogram and correlogram is using. The equation here is used to find out the accurate correspondence of a moving entity of previous frame to another moving entity in the current frame,

$$Mark_i = \cap \sum_{i=1}^{NBC} \sum_{j=1}^{NBP} SIM(i, j) \quad (5)$$

Here NBC is the number of motion blobs in the current input image frame and NBP is the count of the number of motion blob in the previously processed image frame. ∩ denotes that system takes the marker from the previous frame's blob from where the similarity measure value is the minimum and above system's threshold. This is how the tracking occurs. There is exception for the first case that is when the first image frame is processed. For the first time there is no previous frame that's why only a unique marker to each blob is given at this step. This marker is consistently continuing in the later frame which is the sign of successful tracking. In case of occlusion this system handles the occluded entities as a single one and continue to track the group of people until they split. After an occluded object split, system tracks individually each entity. This system may change the marker of tracking but it holds the control over the object which might be useful for security purposes. It is very difficult to resolve the occluding situation in the surveillance system. This is because of camera is missing the occluded entity. The information needed to track the occluded object remains unavailable during an occlusion. Tracking application faces the occlusion problem in different ways. But none of these ways are stand alone to resolve problem if system uses single camera. System with multiple camera arrangement can remove the problem successfully. As our system is based on single camera, the problem is defended by tracking the occluded entity as a whole.

Fig. 2. Processed output of system in occluding situation

It can be seen from Fig 2 that two people were tracked individually in the first four frames. They are marked with separate marker. In the 5th frame an occlusion occurs and the occluded object is tracked with a single marker. In the 6th frame of the figure the occluded object breaks into parts. As soon as the occluded object splits they are being tracked individually with their respective marker again. Thus system holds the control over the moving entities of the scene in the situation where occlusion occurs.

3.7 . Update

To handle gradual illumination change and maintain the background consistently the system updates both the models and background frame adaptively. All of those things can be updated adaptively by following equation,

$$obj(t)=\alpha*obj(t)+(1-\alpha)*obj(t-1) \qquad (6)$$

Where α is the updating constant which determines how slowly or rapidly the updating incorporates new value for respective objects. Value of α may vary for those three objects namely histogram, correlogram and background frame. Here t denotes the current instance of things so $(t-1)$ denotes previous.

IV. EXPERIMENT RESULT & DISCUSSION

We have tested our system on various real time video scene and found satisfactory result. A demonstration of the performance of the system on scattered background and untidy lighting condition over the scene in given in Fig 3.

Fig. 3. Processed frames (1st, 50th, 150th, 198th) of a test video of challenging indoor environment

This result shows that the system is robust under challenging situation with asymmetric lighting condition over the scene and disturbing background. We can see when the entities turn around after an occlusion, system still holds accurate performance. The system holds the total control over the entities in the scene in any situation

A comparison of our system's output and standard CAMSHIFT's tracked output over the video cited in the earlier section is given below in the table. The table shows the displacement error between the ground truth data and CAMSHIFT's result in the second column. Displacement error between ground truth data and developed system's results is showing column titled as Displacement Error in the Table 1. Comparison is shown for 25 frames.

Table 1 shows center location of region of interest (ROI) in each frame from 1 to 25. In the first column given co-ordinates are ground truth data which means actual center location of ROI in image frame. Column named gives the co-ordinates gained from CAMSHIFT and displacement error from ground truth data. Column titled System's Results specifies co-ordinates of center of ROI from developed system's output and displacement error with ground truth. From the comparison it can be seen that, our system is giving much better center of ROI with compare to standard CAMSHIFT results.

Table 1.Comparision of Data

Ground Truth Data (x, y co-ordinates)	CAMSHIFT Results		System's Results	
	(x, y)	Displacement Error	(x, y)	Displacement Error
598, 376	596, 368	8.25	600, 375	2.24
598, 375	572, 361	29.53	596, 375	2.00
593, 376	572, 363	24.70	596, 375	3.16
593, 376	568, 363	28.18	593, 375	1.00
588, 375	567, 364	23.71	593, 375	5.00
581, 375	561, 368	21.19	587, 375	6.00
581, 375	554, 367	28.16	581, 375	0.00
576, 375	552, 363	26.83	581, 375	5.00
572, 375	545, 364	29.15	575, 375	3.00
572, 375	541, 365	32.57	570, 375	2.00
567, 375	540, 365	28.79	570, 375	3.00
566, 375	535, 367	32.02	564, 375	2.00
560, 375	535, 367	26.25	564, 375	4.00
560, 375	530, 369	30.59	558, 375	2.00
554, 375	529, 370	25.50	558, 375	4.00
547, 374	526, 371	21.21	551, 375	4.12
547, 375	523, 369	24.74	545, 374	2.24
541, 375	524, 368	18.38	545, 374	4.12
539, 374	523, 370	16.49	541, 374	2.00
538, 374	520, 370	18.44	537, 374	1.00
536, 373	520, 369	16.49	537, 374	1.41
536, 373	518, 368	18.68	535, 373	1.00
534, 374	518, 368	17.09	535, 373	1.41
534, 373	516, 367	18.97	533, 373	1.00
532, 373	516, 367	17.09	533, 373	1.00

V. CONCLUSION AND FUTURE WORK

In this paper we proposed an appearance based tracking system for tracking human motion from video. Correlogram and histogram based model is used to separate each human and mark them with different marker. The system works efficiently in the indoor environment although it can be appropriated for outdoor environment with a little modification. The performance degradation of the proposed tracking method could have been overlooked in case any rectangular boxes were used as markers of the moving object. But all the moving entities could not be tracked and identified successfully. Our system can maintain the control over the moving entities of the frame even in scattered ambience which is the ultimate goal of any tracking system. The performance of the proposed system may degrade when people under surveillance have dress with similar color or pattern. The system can also be extended in different environmental conditions such as rain, fog, night time etc. The occlusion resolving method could also be added to this system to make the performance better.

REFERENCE

[1] Shireen Y. Elhabian, Khaled M.E1-Sayed and Sumaya H. Ahmed,"Moving object detection in spatial domain using background removal techniques- State-of-Art.", Recent Patents on Computer Schiene, 2008, Vol. 1, No. 1

[2] Prithiviraj banerjee and Somnath Sengupta, " Human motion detection and tracking for video surveillance.", National Conference for communication, 2008, Bombay, India

[3] Jianpeng Zhou and Jack Hoang, "Real time robust human detection and tracking system.", Conference on Face and Gesture Recognition, April, 1998

[4] Pranab Kumar Dhar, DM. Hasibul Hasan, Springerlink, "Real time moving object detection based on masking approach.", International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 5, No. 3, September, 2012

[5] S.S. Beauchemin and J.L Barron, "The computation of optical computing surveys, Vol.27, 1995, pp 433-466

[6] Qi Zhao and Hai Tao, "Object tracking using color correlogram.", 14th International Conference on Computer Communications and Networks, page 263-270

[7] Mart´ı Balcells Capellades, David Doermann, Daniel DeMenthon and Rama Chellappa, "An appearance based approach for human and object tracking.", Conference on Image Processing 2003, ICIP 2003

[8] Jing Huang, S. Ravi Kumar, Mandar Mitra, Wie-Zheng Jhu, Ramin Jabih, "Image indexing using color correlogram.", Conference on Computer Vision and Patternt Recognition, Page 762

[9] Mary J. Comer and Edward J. Delp, "Morphological operations for digital image processing.", Purdue University

[10] Yang Songy, Xiaolin Fengy and Pietro Peron, "Toward detection of human motions.", Computer Vision and Pattertn Recognition, 200

[11] Vlademer nedovik, "Tracking Moving Video Objects using Mean Shift Algorithm.", 6th International Conference on Automatic Face and Gesture Recognition.

[12] Oklahoma cooperative extension service, 4-H youth development, http://www.oces.okstate.edu/washita/uploaded_files/4h_Learni ng_Styles.doc