Human Tracking and Segmentation using Color Space Conversion

Sivabalakrishnan.M[1],Dr.D.Manjula[2]

[1]Research scholar, Department of CSE, College of Engineering, Anna University,
Chennai, TamilNadu, India.

[2]Assistant Professor, Department of CSE, College of Engineering, Anna University,
Chennai, TamilNadu, India.

Abstract

The HSV give more accurate and more robust tracking results compared to grayscale and RGB images. A simple HSV histogram-based color model is used to develop our system. First, a background registration technique is used to construct a reliable background image. The moving object region is then separated from the background region by comparing the current frame with the constructed background image. This paper presents a novel human motion detection algorithm that based regions. This approach first obtains a motion image through the acquisition and segmentation of video sequences. In the situations where object shadows appear in the background region, a pre-processing median filter is applied on the input image to reduce the shadow effect, before major blobs are identified. The second step is generating the set of blobs from detected varied regions in the each image sequence.

Key words: *Median filter, object tracking, background subtraction, rgb2hsv*

1. Introduction

Moving object tracking is one of the challenging task in computer vision problem such as visual surveillance, human computer interactions etc. The act of tracking in video surveillance is becoming an important task especially in monitoring large scale environments such as public and security sensitive areas. In the field of video surveillance, an object of interest would be identified and then monitored or tracked. People are typically the object of interest in video surveillance applications, for example when walking through a secluded or security sensitive area. There is now increasing interest in monitoring people in public areas, e.g. shopping malls. When tracking objects of interest in a wide or public area, additional parameters are required to improve performance such as color of clothing [1], path and velocity of tracked object [2,3] and modeling set colors for tracked persons [4]. To obtain robust tracking of a target, a number of tracking methods are typically employed in order to overcome problems such as occlusion [1, 4, 5] and noise in surveillance videos. Tracking objects is performed in a sequence of video frames and it consists of two main stages: isolation of objects from background in each frames and association of objects in successive frames in to trace them. Background subtraction techniques are mostly used for detection of motion in many real-time vision surveillance applications. In these approaches, difference between the incoming frame and the background image is performed to detect foreground objects. Background subtraction provides the most complete feature data, but is extremely sensitive to dynamic scene changes due to illumination changes and extraneous events. Most researchers are now devoted to develop robust background model in order to prevent some kind of falseness in motion detection caused by scene changes.. The background model is periodically updated by using a combination of pixel-based method and object-based method. . But, unfortunately robust and efficient object tracking is still an open research issue.

To start object tracking, generally the trackers need to be initialized by an outer component [6]. Object tracking in image processing is usually based on reference image of the object, or properties of the objects [7]. Unlike other background maintenance approaches, we design no extra procedure to update background model, instead, a re-initialization of background model method is periodically performed to obtain the newest background scene. In many vision surveillance applications, the moving cast shadows mostly exhibit a challenge for accurate moving targets detection. Moving cast shadows can cause object merging, object shape distortion, and even object losses (due to the shadow cast over another object). For this reason, moving shadows detection is critical for accurate objects detection in vision surveillance applications. In recent years, many algorithms have been proposed in the literatures that deal with shadows. In our present first case, we use HSV histogram based model to make a robust color based tracker system by median filter. Blobs are defined as a group of pixels that belong to a similar object in motion. They have proven to be a better feature cue than points, corners or ridges as they usually have a larger coverage area and total occlusion of the subject is more unlikely to happen. Rossi and Bozzoli [8] successfully used moving blobs to track and count people crossing the field of view of a vertically mounted camera. In a different approach with blobs, Bregler [9] represented each pixel in each motion image by its optical flow

characteristics. These flows are then grouped into blobs that have coherent motion and are characterized by a mixture of multivariate Gaussians. We can find the more details in overall proposed system model section.
The rest of the paper is organized as follows .Section 2 describes Overview of Background subtraction. Section 3 describes proposed system models for object tracking. And then Section 4 shows Experimental Results. Finally the conclusion is given in Section 5.

2. Overview of Background subtraction

Background subtraction is one of the most popular methods for novelty detection in video streams. Background Subtraction, which generates a foreground mask for every frame. This step is simply performed by subtracting the background image from the current frame. When the background view excluding the foreground objects is available, it becomes obvious that the foreground objects can be obtained by comparing the background image with the current video frame .It focuses on two major steps; first, to construct a statistical representation of the background that is representative, robust to the noise and is sensitive to new objects, and second, to build another statistical model called 'foreground' that represents the changes that take place on the scene. By applying this approach to each frame one effectively achieves tracking any moving object a background image can be elegantly used to determine the foreground objects by comparing the input frame with the background image and mark the differences as foreground objects. This technique is commonly known as background subtraction or change detection. It is the most popular approach in video surveillance applications, because it is a computationally efficient technique and it is relatively easy to obtain background images for static surveillance cameras.[10]
In practice, camera noise and regions in which the object has the same color as the background make the separation of foreground objects and background more difficult. Finally, a few post processing filters are presented that can remove obvious errors like small clutter regions. consecutive images is also used for reducing noise in result of the background subtraction. In this process, we assume that moving object is only a human. We then detect moving human region in each image.

3. Proposed system

The proposed algorithm consists of five stages image acquisition, RGB to HSV conversion, BitXOR operation, preprocessing and/ blob identification. Figure 1 shows the process flow of the proposed human motion detection algorithm. Each of these stages will be described in detail .We extract features in the RGB color space. Two feature variables, chromaticity and brightness distortions are used to classify the foreground and background. The color model used here separates the brightness from the chromaticity components. The foreground and background classification is based on the following observation. Image blocks in the background should have little change in their color distortion

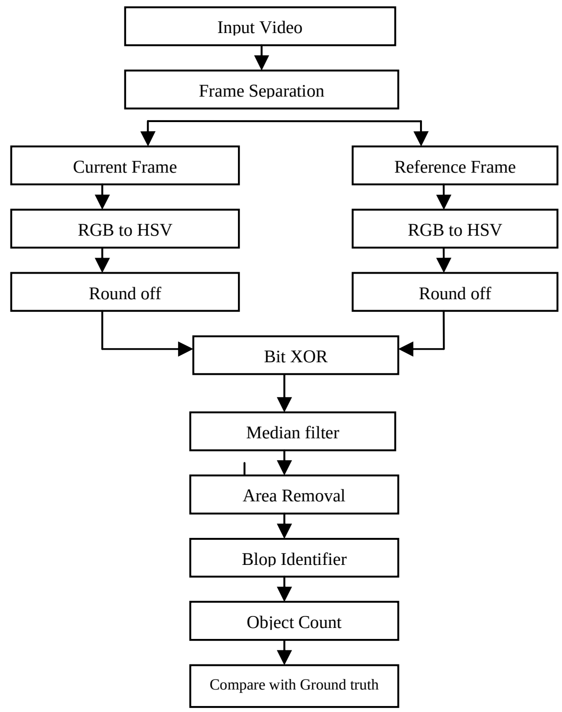


Fig 1 Proposed Tracking System

3.1 Image Acquisition and Segmentation

Image acquisition is a common preliminary step in any motion-based vision application to obtain image frames from a stationary or moving camera, or multiple cameras. Usually, a frame grabber is used to sub sample a sequence of video images at a certain frame rate before the actual processing begins. Generally, video sequences can be of high frame rate (10-30*fps)* or low frame rate (< 10 *fps)*1. It is crucial to take into consideration the type of video system used so that the appropriate segmentation method can be implemented. After the acquisition of image frames, image segmentation can be performed using background subtraction, depending on the frame rate of the video sequences.

3.1.1 Frame segmentation

For high frame rate sequences, the adjacent frame subtraction method is used since the change of motion between consecutive frames is very small. Thus, the difference image, $D(x,y,t)$ between an input frame and the next acquired frame after a fixed time interval c is given . The more widely known background subtraction technique is used for low frame rate sequences where the change of 1 *fps* denotes frames-per-second, the standard measure of frame rate motion is larger. This method eliminates the stationary background, leaving only the desired motion regions. In certain video systems, the acquisition process may be susceptible to erratic changes in illumination, reflection, and noise.

3.2 RGB to HSV Color space conversion

Color vision can be processed using RGB color space or HSV color space. RGB color space describes colors in terms of the amount of red, green, and blue present. HSV color space describes colors in terms of the hue, saturation, and value. In situations where color description plays an integral role, the HSV color model is often preferred over the RGB model. RGB defines color in terms of a combination of primary colors, where as, HSV describes color using more familiar comparisons such as color, vibrancy and brightness. The color spaces that are typically used in video tracking and surveillance are YCbCr [11] and HSV [12].

HSV (M) returns an M-by-3 matrix containing an HSV color map. HSV, by itself, is the same length as the current figure's color map. The colors begin with red, pass through yellow, green, cyan, blue, magenta, and return to red. The map is particularly useful for displaying periodic functions. The mathematical relation between HSV and RGB is given by

$$H = \begin{cases} \theta & \text{if } B \le G \\ 360^0 - \theta & \text{if } B > G \end{cases} \qquad (1)$$

$$\text{Where } \theta = \cos -1 \frac{½ \,(R - G) + (R \_ B)}{[(R-G)^2 + (R-B)(G-B)]^{1/2}}$$

$$S = 1 - [3/(R+G+B)]\,[\min(R,G,B)] \qquad (2)$$

$$V = 1/3\,(R+G+B) \qquad (3)$$

Once the color space conversion is done for both current and background frames, coefficients are then rounded and bit xor mutually.

3.3 Preprocessing

The preprocessing stage consists of two tasks – morphological operations and blob Identification, and they are intended to prepare the motion image for the blob identification stage.

3.3.1. Morphological operations

The following morphological operations are performed on the motion image. The median filter is a nonlinear digital filtering technique, often used to remove noise. Such noise reduction is a typical pre-processing step to improve the results the median filter can eliminate the effect of input noise values with extremely large magnitudes Removal of motion at boundary – Pixels of the motion region that are located along the boundary are eliminated to avoid ambiguity of the region belonging to a possible moving object.

4. Experimental Results

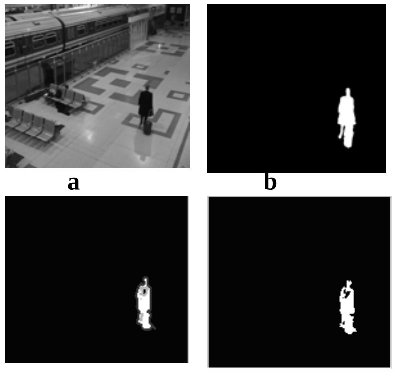
In our experiments we have use PET2006 Database, which consists of video sequences of human movements in various gait poses, non-human object motions and a combination of human and object movements, constructed from our motion capture system

In performance evaluation, each pixel in a background subtraction method's classification was determined to be, true positive for a correctly classified foreground pixel, false positive for a background pixel that was incorrectly classified as foreground, true negative for a correctly classified background pixel, and false negative for a foreground pixel that was incorrectly classified as background. the different methods can be evaluated by the calculation of TP, TN, FP, FN, where TP, TN, FP, FN means the number of true positive, true negative, false positive and false negative, respectively. After every pixel had been classified into one of those four groups, the sensitivity, the specificity was calculated.

Sensitivity and specificity are statistical measures of the performance of a binary classification test**.** Sensitivity (also called recall rate in some fields) measures the proportion of actual positives which are correctly identified .Specificity measures the proportion of negatives which are correctly identified. These two measures are closely related to the concepts of type I and type II errors. sensitivity is defined in equation 4, specificity is defined in equation 5

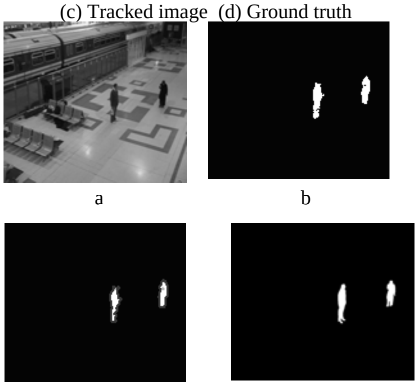
$$\text{Sensitivity} = TP / (TP + FN) \qquad (4)$$
$$\text{Specificity} = TN / (FP + TN) \qquad (5)$$

The sensitivity measures the proportion of actual positives which are correctly identified. The specificity measures the proportion of negatives which are correctly identified. These metrics are a measure of the accuracy or the correctness of the tracking methodology that is being evaluated with respect to a reference point, known as the ground truth. Ground truth can be defined as reference or baseline data for determining the actual path of a tracked object. A correct track is considered true when the tracked point is within the boundary region of the tracked object.



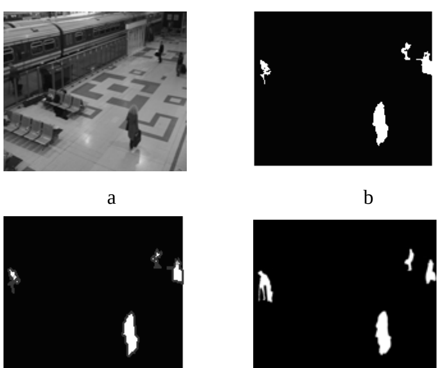
a b c d

Fig 2 (a) Original image (b) Foreground image
(c) Tracked image (d) Ground truth



a b c d

Fig 3 (a) Original image (b) Foreground image
(c) Tracked image (d) Ground truth



a b c d

Fig 4 (a) Original image (b) Foreground image
(c) Tracked image (d) Ground truth

Table 1 shows the quantitative results of various frames. In a preliminary motion-free test, the system took 616 seconds to process 1,000 image frames on an Intel Pentium IV 2.63Ghz, 504MB RAM system, which is approximately 1.62 *fps*. MATLAB was used as the implementation tool.

Table 1: Quantitative Evaluation of test Results

S.I .N O	No of objects	TP	TN	FP	FN	Sensitivity	Specificity
1	1	6716	407169	821	821	0.8911	0.9980
2	2	5143	409147	614	614	0.8933	0.9985
3	4	10060	401764	1365	1365	0.8805	0.9966

5. Conclusions

In this paper, we proposed an accurate and robust object extraction scheme for a dynamic environment. We built a background model and extract classification features in RGB color space. To deal with the challenges of object extraction in dynamic environment, we fused high-level knowledge and low-level features and developed a fuzzy logic inference system. The results on several sequences show that this algorithm is efficient and robust for the dynamic environment with new objects in it. We are currently working on making the prediction more accurate and creating a scheme to recover missing moving parts using known results from feature-based classification. Also, we intend to study the impact of the accuracy of results on the performance of future activity modeling and analysis. When object are closer the segmented object gets failed and it was detected as a single object, this error will be rectified by using fuzzy logic in the future work.

References

[1] Bird, N.D., et al. Detection of loitering Individuals in public transportation areas in IEEE Transactions on Intelligent Transportation Systems. 2005.
[2] Bodor, R., B. Jackson, and N.P. Papanikolopoulos. Vision-Based Human Tracking and Activity Recognition. in 11th Mediterranean Conference on Control and Automation. 2003.
[3] Niu, W., et al. Human Activity detection andrecognition for video surveillance. In IEEE International Conference on Multimedia and Expo 2004.
[4] Iocchi, L. and R.C. Bollees. Integrated Plan-ViewTracking and Color-based person Models for Multiple People Tracking. in International Conference on Image Processing 2005.
[5] Li, J., C.S. Chua, and Y.K. Ho. Color based multiple people tracking. in 7th International Conference on Control Automation, Robotics andVision 2002. ICARCV 2002.
[6] C. Yunqiang and R. Yong, "Real time object tracking invideo sequences", Signals and Communications Technologies, Interactive Video, 2006, Part II: pp. 67 – 88
[7] K. Nummiaro, E. Koller-Meier, and L. Van Gool, "A color-based particle filter", In Proc. First International Workshop on Generative-Model-Based Vision, 2002, pp. 53– 60

[8] Rossi, M., and Bozzoli, A. 1994. Tracking and Counting People. In 1st Int. Conf. on ImageProcessing, 212-216.

[9] Bregler, C. 1997. Learning and Recognizing Human Dynamics in Video Sequences. In Proc. of IEEE CS Conf. on Comp. Vision and Patt. Recognition,568 - 574.

[10]Sivabalakrishnan.M,Dr.D.Manjula“Adaptive Background Subtraction in Dynamic Environment using Fuzzy Logic ”International journal on Computer Science and Engineering,Vol2,270-273.

[11] Chen, Y.-J., et al. The Implementation of a Standalone Video Tracking and Analysis System for Animal Behavior Measurement in Morris WaterMaze. In 27th Annual International Conference on Engineering in Medicine and Biology Society 2005 (IEEE-EMBS 2005). Shanghai.

[12] Stern, H. and B. Efros. Adaptive Color Space Switching for Face Tracking in Multi-Colored Lighting Environments in Fifth IEEE Conference on Automatic Face and Gesture Recognition 2002. Washington, D.C

M.Sivabalakrishnan is a Research Scholar,Department of CSE,College of Engineering,Anna University,Chennai 600 025,India.He received the Master’s degree from P.S.N.A College of Engineering, Anna university, Chennai., in 2004 . His main research interests include multimedia and image processing.

Dr.D.Manjula received the Master’s and Ph.D degree from College of Engineering, Anna University, Chennai, India in 1987and 2004 respectively. She is currently working as Assistant Professor in the Department of Computer Science and Engineering, Anna University, Chennai. Her research interests include NLP, Text Mining, Parallel processing ,Grid computing, Database and multimedia.