IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified

Vol. 5, Issue 2, February 2017

# Moving Object Detection and Tracking

**C. Ranjeeth Kumar[1], S.S. Sugantha Mallika[2], J. Sree Ranjaane[3]**

Assistant Professor (Sr.Gr), Department of IT, Sri Ramakrishna Engineering College, Coimbatore, India[1]

Assistant Professor, Department of IT, Sri Ramakrishna Engineering College, Coimbatore, India[2]

Student, Department of IT, Sri Ramakrishna Engineering College, Coimbatore, India[3]

**Abstract**: The detection of moving objects in videos is very important in many video processing applications and background modeling is often an indispensable process to achieve this goal. Most of the traditional background modeling methods utilize color or texture information. However, color information is sensitive to illumination variations and texture information cannot be utilized to separate smooth foreground from smooth background in most cases. A new integration framework of texture and color information for background modeling, in which the foreground decision equation includes three parts (one part for color information, one part for texture information and the left part for the integration of color and texture information). This framework is able to combine the advantages of texture and color features while inhibiting their disadvantages as well. A block based method to accelerate the background modeling. Specifically, in the texture information modeling process, a single histogram model is established for each block whose bins indicate the occurrence probabilities of different patterns, which is different from the traditional multi-histogram model for block-based background modeling, and then dominant background patterns are selected to calculate the background likelihood of new coming blocks. Dynamic background and multimodal problems can be handled through this technique.

**Keywords:** Integrated information, Moving objects, Object detection, Image texture.

## I. INTRODUCTION

The use of video is becoming prevalent in many applications such as monitoring of traffic, detection of pedestrians, identification of anomalous behavior in a parking lot or near an ATM, etc. A single image provides a snapshot of a scene, the different frames of a video taken over time registers the dynamics in the scene, making it possible to capture motion in the sequence.

A key task in mining video data is the detection and tracking of moving objects, such as people and vehicles, through the video frames. Motion is very important in making objects easy to recognize as soon as they move, even if they are inconspicuous when still. This allows us to model their interactions and to detect unusual events. The detection and tracking of moving objects is a task which must be performed accurately and robustly to minimize false alarms and in real-time to enable corrective action.

Object recognition is a process for identifying a specific object in a digital video or image. Object recognition algorithms rely on matching, learning, or pattern recognition algorithms using appearance-based or feature based techniques. Task of finding and identifying objects in an image or video sequence is easy. Humans recognize a multitude of objects in images with little effort, despite the fact that the image of the objects may vary somewhat in different viewpoints, in many different sizes and scales or even when they are translated or rotated.
Objects can be recognized when they are partially obstructed from view. This task is still a challenge for computer vision systems. Many approaches to the task have been implemented over multiple decades.
Background subtraction, also known as Foreground Detection, is a technique in the fields of image processing and computer vision wherein an image's foreground is extracted for further processing (object recognition etc.). Generally an image's regions of interest are objects (humans, cars, text etc.) in its foreground. After the stage of image pre-processing (which may include image denoising, post processing like morphology etc.) object localisation is required which may make use of this technique. Background subtraction is a widely used approach for detecting moving objects in videos from static cameras.

Detection of moving objects in video can be difficult for several reasons. We need to account for possible motion of the camera, changes in illumination of a scene, objects such as waving trees, objects that come to a stop and move again such as vehicles at a traffic light, etc. Once the moving objects have been identified, tracking them through the video sequence can also be difficult, especially when the objects being tracked are occluded by buildings or move in and out of the frame due to the motion of the camera.

A flexible tracking pipeline that allows us to investigate different combinations of foreground extraction, feature extraction and motion correspondence algorithms. Parts of this tracking pipeline can also be applied.

IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified

Vol. 5, Issue 2, February 2017

In foreground extraction we have explored applications of background subtraction and salient region extraction. Our background subtraction research includes the investigation of the effectiveness of popular background subtraction techniques and the development of a new technique for background subtraction with foreground validation.

Foreground object detection and segmentation from a video stream is one of the essential tasks in video processing, understanding, and object-based video encoding (e.g., MPEG4). A commonly used approach to extract foreground objects from the image sequence is through background suppression or background subtraction when the video is grabbed from a stationary camera. These techniques have been widely used in real-time video processing. However, the task becomes difficult when the background contains shadows and moving objects, e.g., wavering tree branches and moving escalators, and undergoes various changes, such as illumination changes and moved objects.

Many methods have been proposed for real-time foreground object detection from video sequences.

Video tracking is the process of locating a moving object (or multiple objects) over time using a camera. It has a variety of uses, some of which are: human-computer interaction, security and surveillance, video communication and compression, augmented reality, traffic control, medical imaging and video editing. Video tracking can be a time consuming process due to the amount of data that is contained in video. Adding further to the complexity is the possible need to use object recognition techniques for tracking, a challenging problem in its own right. The following are some common target representation and localization algorithms:

- Kernel-based tracking (mean-shift tracking): an iterative localization procedure based on the maximization of a similarity measure (Bhattacharyya coefficient).
- (e.g. active contours or Condensation algorithm). Contour tracking methods iteratively evolve an initial contour initialized from the previous frame to its new position in the current frame. This approach to contour tracking directly evolves the contour by minimizing the contour energy using gradient descent.

The objective of video tracking is to associate target objects in consecutive video frames. The association can be especially difficult when the objects are moving fast relative to the frame rate. Another situation that increases the complexity of the problem is when the tracked object changes orientation over time. To perform video tracking an algorithm analyzes sequential video frames and outputs the movement of targets between the frames. There are a variety of algorithms, each having strengths and weaknesses.

## II. PROPOSED SYSTEM

Moving object detection plays an important role in many video processing applications such as object tracking, categorization, re-identification and video condensation

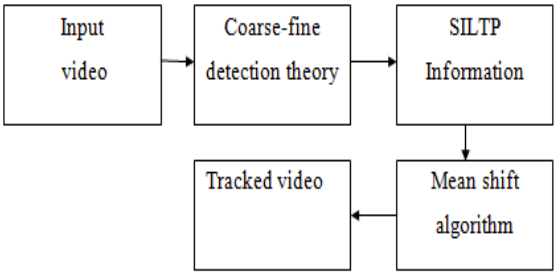


**Fig 2.1: Block diagram of proposed system**

It often serves as preprocessing for higher level video analyses and its performance directly affects the performance of the subsequent applications. For object tracking, if a moving object is detected as two or two moving objects are detected as one, the tracking result may be incorrect. For object categorization, in completeor adhesive detection of moving objects may lead to wrong categorization, and it is the same case for object re-identification. For video condensation, object tracking is also an indispensable part. It is not the desired result if the head and legs of one person appear at different time in the condensed video. Ideally, a detection method should detect each moving object separately without breaking. For detecting the video coarse to fine detection theory is used. For color information SILTP algorithm is used.

### A. FOREGROUND AND BACKGROUND MODEL CONSTRUCTION

In foreground and background model construction the input video is divided into various frames. To represent observed image regions, we first apply a new binary descriptor. Background models with respect to each pixel pcan be efficiently built. Since these background instances are computed from the observed backgrounds of individual frames, they can represent the most recent background changes.The Foreground Detector System object compares a color or grayscale video frame to a background model to determine whether individual pixels are part of the background or the foreground. It then computes a foreground mask. By using background subtraction, we can detect foreground objects in an image taken from a stationary camera. More specifically, our background models are built by a temporal series of background instances, which contain observed backgrounds.

To avoid effects of illumination changes and achieve real-time performance, each instance is represented by a binary descriptor. When a new frame comes, instances are extracted and represented by binary descriptors computed from image regions at first. The label of each instance is

IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified

Vol. 5, Issue 2, February 2017

decided by the coarse level detection theory based on both background and foreground models. Each instance is computed from a region, the fine level detection theory is applied to identify the label of each pixel.

**B. COARSE TO FINE DETECTION THEORY**

Fixed thresholds are hard to be applied to different videos. It is also very hard to change threshold automatically with respect to different kinds of surveillance videos. Thus, some approaches such as aim to modify thresholds based on information of current frames. However, due to variant background changes, it is still hard to retrieve good thresholds, which can be applied to different videos. Moreover, not only the distance threshold is required, but also an additional threshold of the number of similar background samples with respect to the incoming sample, i.e. the amount threshold, is required. Only if sufficient background samples match the incoming sample, the incoming sample is considered as a background.

To reduce the number of thresholds used in most nonparametric methods, a new coarse-to-fine detection theory algorithm to identify the labels of each region and pixel. The foreground extraction problem as two binary hypothesis-testing problems in the region and pixel levels, respectively. Identify the labels of each instance by the detection theory in the coarse level. As a result, decide the labels of each region and pixel.

**COARSE-LEVEL DETECTION THEORY**

Coarse level detection is meant for background elimination.For finding an instance whether it belongs to either a background or a foreground instance, identifying the label of naturally forms a binary hypothesis-testing problem. Here, we consider two hypotheses. For deciding an instance to be a foreground or background a hypothesis testing problem were used.

$$\Lambda_H = \frac{P(I_p(t')|H_0)}{P(I_p(t')|H_1)} \underset{H_1}{\overset{H_0}{\gtrless}} \frac{P(H_1)}{P(H_0)}.$$

To solve the binary hypothesis-testing problem in the coarse level, we propose using the Bayesian maximum a posterior (MAP) detection theory to decide the label of each instance. The instance Ip(t') can be classified as a foreground instance.Only half of the background gets eliminated by this method. The output of coarse is given to fine level detection theory.

**FINE LEVEL DETECTION THEORY**

Fine level detection theory eliminates the background on the basis of its pixels.Each instance represents an image region, all of the pixels in the region are then considered to be the same label. As a result, only rough shapes of foreground objects can be retrieved. To further extract detailed shapes of foreground objects, i.e. identifying if a pixel is a foreground pixel or not, we employ the fine level detection theory to decide the label of the pixel.

$$\begin{cases} G_0: \text{ p(t') is a background pixel} \\ G_1: \text{ p(t') is a foreground pixel} \end{cases}$$

The MAP detection theory in the fine level to decide the label of **p**(t') based on labels of instances containing **p**(t'). The instances containing **p** (t') form an union.

$$P(G_0 \mid p(t')) \underset{G_1}{\overset{G_0}{\gtrless}} P(G_1 \mid p(t')).$$

By using the coarse to fine detection theory, the detailed shape of the objects can be retrieved. In fine level detection theory the rough shape of the foreground can only be retrieved.

**C. SCALE INVARIENT LOCAL TERNARY PATTERN (SILTP)**

A scale invariant local ternary pattern operator is effective for handling illumination variations, especially for moving soft shadows. A pattern kernel density estimation technique effectively model the probability distribution of local patterns in the pixel process, which utilizes only one single LBP-like pattern instead of histogram as feature. Multimodal background models with the multiscale fusion scheme for handling complex dynamic backgrounds. Unlike LBP, it does not threshold the pixels into 0 and 1, rather it uses a threshold constant to threshold pixels into three values. Considering k as the threshold constant, c as the value of the centre pixel, a neighbouring pixel p.

$$\begin{cases} 1, & \text{if } p > c + k \\ 0, & \text{if } p > c - k \text{ and } p < c + k \\ -1 & \text{if } p < c - k \end{cases}$$

In this way, each threshold pixel has one of the three values. Neighboring pixels are combined after thresholding into a ternary pattern. A scale invariant local ternary operator for handling illumination variations. By combining texture and colour, the detection is improved.

Given any pixel location, SILTP encodes as

$$\text{SILTP}\ (x_c, y_c) = \bigoplus_{K=0}^{N-1} S\ (I_C, I_K)$$

Where Ic is the gray intensity value of the centre pixel, Ik are that of its N neighbourhood pixels. Concatenation

IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified

Vol. 5, Issue 2, February 2017

function used. In this way, each threshold pixel has one of the three values. Neighboring pixels are combined after thresholding into a ternary pattern. Computing a histogram of these ternary values will result in a large range, so the ternary pattern is split into two binary patterns.

$$s_\tau(I_c, I_k) = \begin{cases} 01, \ if \ I_k > (1+\tau)I_c, \\ 10, \ if \ I_k < (1-\tau)I_c, \\ 00, \ otherwise. \end{cases}$$

Each comparison results with one of the three values, SILTP encodes it with two bits. Extensive experiments on complex scenes have to run to prove that one single local texture pattern instead of region histogram is really enough for the background subtraction task. The pattern is cross-calculated in RGB color channels with the Scale Invariant Local Ternary Pattern operator. The descriptor shows an excellent performance in texture regions, in flat regions. The SILTP operator performs perfectly, with only two patterns being different between the two black blocks and if the neighbouring pixel is similar to the centre pixel then it is marked as 00. The block based approach is followed. Each small block may belong to several big blocks. After background modeling, each big block will have histogram of background.

**PROCEDURE**

- Divide the examined window into cells (e.g. 16x16 pixels for each cell).
- For each pixel in a cell, compare the pixel to each of its 8 neighbors (on its left-top, left-middle, left-bottom, right-top, etc.).
- Follow the pixels along a circle, i.e. clockwise or counter-clockwise where the center pixel's value is greater than the neighbor's value, write "1". Otherwise, write "0". This gives an 8-digit binary number (which is usually converted to decimal for convenience).
- Compute the histogram, over the cell, of the frequency of each "number" occurring (i.e., each combination of which pixels are smaller and which are greater than the center).
- Optionally normalize the histogram.
- Concatenate (normalized) histograms of all cells. This gives the feature vector for the window.
- Several blocks are concatenated.

The feature vector can now be processed using the Support vector machine or some other machine-learning algorithm to classify images. Such classifiers can be used for face recognition or texture analysis.

**ADVANTAGES**

- It is computationally efficient which causes only one comparison.
- Integrating texture and color information with a greater efficiency.
- The SILTP operator is robust to local image noises within a range.
- The scale invariance property makes SILTP robust to illumination changes.

**D. MEAN SHIFT ALGORITHM**

Mean shift algorithm is generally used in tracking visible data, and can be implemented on either colour or monochromatic image sequences**.** The Mean Shift algorithm (MS) was first used as a gradient decent technique used in pattern recognition. It was used for feature space image segmentation followed by object tracking. It is also known as kernel based object tracking, due to the use of two kernels that are applied to the target before the centroiding part of the algorithm.

The mean shift algorithm can be used for visual tracking. The simplest such algorithm would create a confidence map in the new image based on the color histogram of the object in the previous image, and use mean shift to find the peak of a confidence map near the object's old position. The confidence map is a probability density function on the new image, assigning each pixel of the new image a probability, which is the probability of the pixel color occurring in the object in the previous image.
The algorithm can be viewed as a procedure performed at every frame, and calculate the centre of an object. The object the centroid is performed on is not just a sub-image containing the object of interest, but a probability densityfunction of that object. Here is the general form of the algorithm.

**PROCEDURE**

- Given an initial (or previous) location and object size, create sub-image containing the original object's location, larger than the object's size.
- Create a kernel based on object size to weight the centre pixels more heavily.
- Create a kernel based on the pixel intensity range for the object in question.
- Apply both kernels to the sub-image, creating a probability density function of the likelihood a pixel belongs to the object of interest.
- Sum all pixels or perform a moment calculation on the distribution, finding the centre location.
- Repeat this, using the location found last until convergence is met or a maximum iteration count is exceeded.

## III. ADVANTAGES

- Mean shift is an application-independent tool suitable for real data analysis.
- Does not assume any predefined shape on data clusters.
- It is capable of handling arbitrary feature spaces.
- The procedure relies on choice of a single parameter: bandwidth.

IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified

Vol. 5, Issue 2, February 2017

**SCREEN SHOTS**

Fig 3.1: Input video

Fig 3.2: Construction of Frames

Fig 3.3: Background and Foreground Model Construction

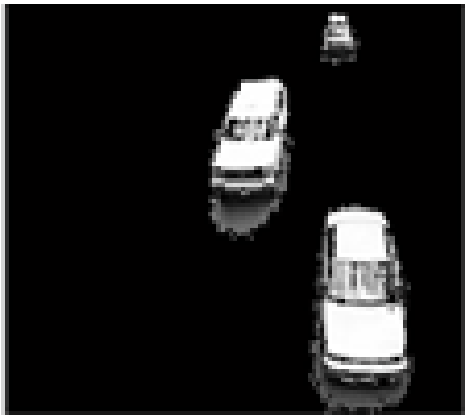

Fig 3.4: Background subtraction

Fig 3.5: SILTP color information

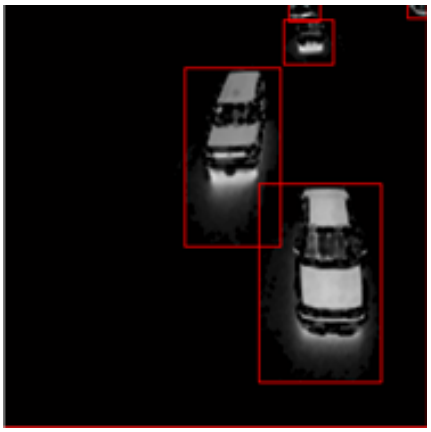

Fig 3.6: Mean shift tracking

## IV. CONCLUSION AND FUTURE ENHANCEMENT

A block-based model with single SILTP histogram has been proposed and is able to handle dynamic background and multimodal problems. The SILTP information and color information have been integrated for much more effective detection of moving objects than separately applied. A new quality measure is proposed for evaluating the performance of our method on various challenging videos, and the result is quite outstanding compared with the other state-of-the-art methods. The memory consumption is low while the processing speed can be super-real-time.

IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified

Vol. 5, Issue 2, February 2017

The future work of our project will be extracting the one particular moving object by giving its feature. The optimized algorithms can be used to make this system robust and reliable.

## REFERENCES

[1] C. R. Wren, A. Azarbayejani, T. Darrell, A. P. Pentland, "Pfinder: Real-time tracking of the human body", IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 7, pp. 780-785, Jul. 1997.
[2] C. Stauffer, W. E. L. Grimson, "Adaptive background mixture models for real-time tracking", Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2, Jun. 1999.
[3] M. Harville, "A framework for high-level feedback to adaptive per-pixel mixture-of-Gaussian background models" in Computer Vision, Berlin, Germany:Springer-Verlag, pp. 543-560, 2002.
[4] D.-S. Lee, J. J. Hull, B. Erol, "A Bayesian framework for Gaussian mixture background modeling", Proc. Int. Conf. Image Process. (ICIP), vol. 3, pp. III-973-III-976, Sep. 2003.
[5] A.Elgammal, R. Duraiswami, D. Harwood, L. S. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance", Proc. IEEE, vol. 90, no. 7, pp. 1151-1163, Jul. 2002.
[6] L. Li, W. Huang, I. Y. H. Gu, Q. Tian, "Foreground object detection from videos containing complex background", Proc. 11th ACM Int. Conf. Multimedia, pp. 2-10, 2003.
[7] K. Kim, T. H. Chalidabhongse, D. Harwood, L. Davis, "Background modeling and subtraction by codebook construction", Proc. Int. Conf. Image Process., vol. 5, pp. 3061-3064, Oct. 2004.
[8] K. Kim, T. H. Chalidabhongse, D. Harwood, L. Davis, "Real-time foreground-background segmentation using codebook model", Real-Time Imag., vol. 11, no. 3, pp. 172-185, 2005.
[9] A.Monnet, A. Mittal, N. Paragios, V. Ramesh, "Background modeling and subtraction of dynamic scenes", Proc. 9th IEEE Int. Conf. Comput. Vis., pp. 1305-1312, Oct. 2003.
[10] M. Heikkila, M. Pietikainen, "A texture-based method for modeling the background and detecting moving objects", IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 4, pp. 657-662, Apr. 2006.
[11] S. Liao, G. Zhao, V. Kellokumpu, M. Pietikainen, S. Z. Li, "Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes", Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 1301-1306, Jun. 2010.
[12] J. Yao, J.-M. Odobez, "Multi-layer background subtraction based on color and texture", Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 1-8, Jun. 2007.
[13] Z. Zhang, C. Wang, B. Xiao, S. Liu, W. Zhou, "Multi-scale fusion of texture and color for background modeling", Proc. IEEE 9th Int. Conf. Adv. Video Signal-Based Surveill. (AVSS), pp. 154-159, Sep. 2012.
[14] E. Learned-Miller, M. Narayana, A. Hanson, "Background modeling using adaptive pixelwise kernel variances in a hybrid feature space", Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 2104-2111, Jun. 2012.
[15] T. Matsuyama, T. Ohya, H. Habe, "Background subtraction for non-stationary scenes", Proc. Asian Conf. Comput. Vis., pp. 662-667, 2000.
[16] M. Mason, Z. Duric, "Using histograms to detect and track objects in color video", Proc. 30th Appl. Imag. Pattern Recognit. Workshop (AIPR), pp. 154-159, Oct. 2001.


DOI 10.17148/IJIREEICE.2017.5215