

An Adaptive Threshold Method of Temporal Difference for Detection of Moving Object and Recognition using Hidden Markov Model

Saurabh Dumbre

*Department of Computer Engineering
Sinhagad College of Engineering, Pune, Maharashtra, India*

Mrudul Arkadi

*Department of Computer Engineering
Sinhagad College of Engineering, Pune, Maharashtra, India*

Aniket Ghorpade

*Department of Computer Engineering
Sinhagad College of Engineering, Pune, Maharashtra, India*

Avinash More

*Department of Computer Engineering
Sinhagad College of Engineering, Pune, Maharashtra, India*

Abstract - In this paper, a method for detection and recognition of moving objects is proposed. Here, a moving object is detected using the Temporal Difference Method (TDM) and the object is recognized by the Hidden Markov Model (HMM). Initially, the pixel value in the image domain is converted to amount of energy change in entropy domain by use of concept of entropy. The temporal difference method detects the region of moving objects in complex images to address changing environments. The feature vectors of the detected mask image are obtained by using the Discrete Wavelet Transform (DWT). The hidden Markov model accurately recognizes the moving objects.

Keywords – Moving object Detection and Recognition, Entropy, Temporal Difference Method, Discrete Wavelet Transform, Hidden Markov Model.

I. INTRODUCTION

The Video Surveillance System has come into day-to-day life gradually and has been widely applied to many places, such as supermarkets, banks, schools, museums, and some others. But the traditional Video Surveillance which is operated by people has been replaced by the intelligent Video Surveillance. The intelligent Video Surveillance rarely needs people to operate, and it can take the place of people's eyes and release human beings from the heavy and boring work. This system mainly deals with the image sequence captured from cameras and then detects and tracks the moving objects automatically. When it finds something wrong, the intelligent Video Surveillance will alarm automatically. Therefore, it greatly reduces one's work and avoids the loss caused by the surveillance. A robust video monitoring system should be capable of dealing with movement through change of background, such as weather, lighting, shadow and some other small unimportant objects. Traditional approaches typically fail in these general situations. Our goal is to find a simple method to realize the real-time detection of the moving objects, so as to apply it to the intelligent surveillance system. At present, there are three conventional approaches to moving objects detection: temporal differencing, background subtraction, and optical flow [4].

In many surveillance applications, background subtraction has been used extensively to detect the foreground regions however, a common problem of background subtraction is that it requires a long time for estimating the background models. Temporal differencing is more effective in the case that the moving objects are obviously different from the background and moving relatively fast. Temporal differencing could rapidly detect the possible

moving objects and be adapted to the changing lightening [6]. Optical flow provides all motion information. But optical flow computation methods are too complex to use in real-time applications if without special hardware.

The present paper proposes a new moving object detection and recognition method that uses the temporal difference method along with the hidden Markov model. In the first step, we use the concept of entropy and then perform the transformation of the gray-value domain of the currently observed image into the Clausius entropy domain. Secondly, the temporal difference method, here quickly detects the region of moving objects in a complex image to address the continuous variation in the environmental conditions. The discrete wavelet transform which is the third step is used to extract the feature vectors from the detected mask image of the moving object. In the last step, we accurately recognize the moving objects where we use the hidden Markov model with the discrete wavelet descriptor as the feature vectors.

II. PROPOSED ALGORITHM

2.1 Detection Algorithm for Moving Object

In this section we are going to detect the moving object in video sequence using Temporal Difference method. This procedure is divided into two steps. First we will calculate the Clausius Entropy of the image. The entropy of an image can be determined approximately from the histogram of the image. The histogram shows the different grey level probabilities in the image. Second, the coarse region of the moving object is detected by calculating the entropy difference between consecutive frames using Temporal Difference.

A. Estimation of Clausius Entropy

Entropy is a function of a quantity of heat in a system that is capable of doing work. Under maximum entropy, there is a minimum amount of energy available for doing work and oppositely, under minimum entropy, there is a maximum amount of energy available for doing work. Entropy S is not defined directly, but rather by an equation relating the change in entropy of the system to the change in heat of the system [8]. A change in entropy (ΔS) is defined by the equation :

$$\Delta S = \frac{\Delta Q}{T}$$

where ΔQ is the amount of heat absorbed in an isothermal and reversible process in which a system goes from one state to another and T is the absolute temperature at which the process is occurring. If the temperature of the system is not constant, then this relationship is represented by a differential equation:

$$dS = \frac{dQ}{T}$$

To understand what this equation means, suppose that the temperature T can be expressed as a function $T(Q)$ of the heat Q . The total change in entropy as the heat-level varies is:

$$\Delta S = \int_A \frac{1}{T(Q)} dQ$$

where A is the set defining the range of heat values in the system [8,9]. For computation of the Clausius entropy for the pixel value of each frame in the image sequence three necessary concepts are: the system or field F , energy or heat Q , and temperature T . First, we define the system or field F as an input video I composed of the gray scale values of color images I_t . Each pixel in the frame image I_t of a video has a $w \times w$ rectangular neighborhood or window.

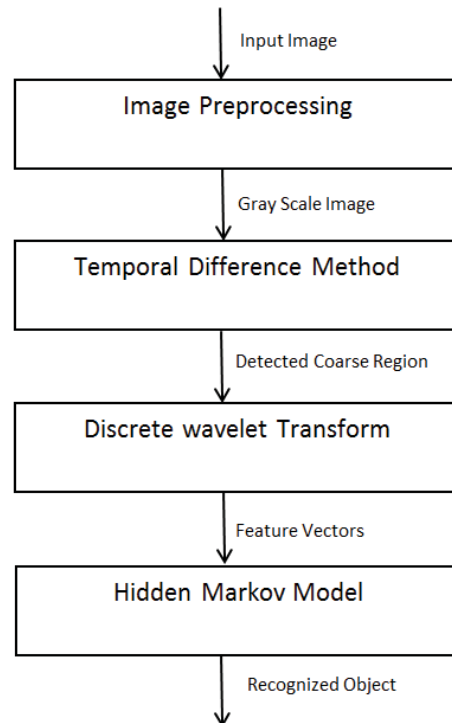


FIGURE 1 : BLOCK DIAGRAM OF THE SYSTEM

Here each pixel, which consists of the window of every frame, takes energy from the former input image and emits it to the next output image. If the energy absorbed from the input image is greater than that emitted to the next frame, then the temperature is increasing, and the temperature is decreasing if it is less. Second, we consider the energy of a given frame. We can define the absorbed energy from the input image follows:

$$\Delta Q_k^{(P)} = \sum_{\text{half of pixels in window}} w_k (X_{ki}^{(P)} - M_{ki}^{(P)})^2$$

Where X_{ki} the color value of k^{th} is channel at the i^{th} pixel in the window, $M_{ki}^{(P)}$ is the mean of all the color values of the channel for pixels and w_k is the weight function for each channel. Further to adapt the system to time, we have to update the mean value $M_{ki}^{(P)}$ for each image frame as follows :

$$M_{ki}^{(P)} = (1 - \lambda) M_{ki}^{(P-1)} + \lambda X_{ki}^{(P)}$$

Where λ is the learning factor for adapting the current mean. The value of learning factor has a range of 0 to 1. In general, the greater the temperature difference between two objects, the more rapid the energy movement is. Similarly, we define that the amount of energy is increasing according to the difference between the color value and mean value is more and more large.

Third, we define the absolute temperature of the system. At the micro level, the temperature can be defined as the average energy of each particle in the system. Hence, if we take κ as the proportional constant between heat and temperature, then the change of temperature in thermodynamic system can be defined as follows :

$$\Delta T = \kappa \frac{\Delta Q}{n}$$

where n is the total number of particles belonging to some object. Hence we can define the temperature in each frame as :

$$T^{(t)} = (1 - \rho)T_E^{(t-1)} + \kappa \frac{\Delta Q}{n}$$

where n is the constant proportional to the amount of heat loss for every frame. The above definition satisfies the rule of the heat system : the greater the difference in temperature between two objects, the more rapid the energy movement. Further, the amount of energy emissions for each frame may be given as

$$\Delta Q_L = \rho \frac{n \cdot \Delta T}{n}$$

Fourth, we compute the total entropy variation for each pixel (x,y) in the t^{th} frame by the sum of entropy variations for each channel :

$$R_{t+1}(x,y) = \sum_{i=1}^3 w_i f_i(x,y)$$

Finally, the Clausius entropy can be used to detect moving regions by making use of the difference of their entropy variations in consecutive frames in a video sequence [8,9].

B. Detection of moving object regions

The temporal differencing method has been used to quickly detect coarse regions of multiple moving objects. Temporal difference is also called frame difference. In the continuous image sequence especially in two or three consecutive video frames, temporal difference is adopted based on pixel and threshold to extract motion areas. It can be easily adapted to static environments, and also it can accurately indicate an initial coarse area of moving object [3,4,6].

In the condition of small changes in lighting, if the absolute difference is smaller than the threshold, then we consider it as background, otherwise foreground. We mark the pixels which belong to foreground. Then we can use the marked areas to extract the moving objects easily.

Let (x,y) and $I_{t+1}(x,y)$ represent the Clausius entropy value at the pixel position (x,y) and at times t and $t+1$ for the video image sequence I with the range $[0,255]$. To detect slow-moving or stationary objects, we can use the adaptive threshold value to compute the temporal difference image $\Delta I_{t+1}(x,y)$, shown by the following equation:

$$\Delta I_{t+1}(x,y) = \begin{cases} 1 & \text{if } |I_{t+1}(x,y) - I_t(x,y)| > T_d(x,y) \\ 0 & \text{otherwise} \end{cases}$$

Where $T_d(x,y)$ is the threshold value estimated from the distribution of entropy differences [4].

In this case, because the system is used for outdoor as well as indoor environments, we need to adapt the threshold value to dynamic changes in the environment, such as changes in global illumination and long-term background updates. Hence, to obtain the adaptive threshold, we first construct a histogram of differences between two entropy images for each pixel. That is, the probability density function of differences $R_{t+1}(x,y)$ is given by the

following form :

$$R_{t+1}(x,y) = \sum_{i=1}^3 w_i f_i(x,y)$$

where $f(x,y)$ denotes respectively the probability density function of pixel value differences for the background, the shadow, and moving objects. Here the first term represents the distribution of absolute differences between pixels in the background, the second term represents the distribution of pixel value differences for the shadow, and the third term represents the distribution of pixel value differences for moving objects.

Hence we, note that the optimal threshold occurs in the middle point where the distributions for the shadow and moving objects meet. Thus, after we identify the mode for each of the three distributions, we take the threshold value as the entropy difference value at the midpoint between the second mode and third mode.

Finally, we detect the coarse region of moving objects by using the following steps: (1) Prepare the temporal tracking mask image for the representative region of the moving object; (2) Compute the difference in the pixel values between two consecutive frames in the image sequence; (3) Compare the computed difference with the threshold and if this is greater than then assign 1 to the pixel value for its position in the temporal tracking mask image, otherwise 0; (4) Apply morphological operations such as erosion and dilation to the region of coarse detection to remove noise or to restore some missing pixels of moving objects. We then use the coarse detected mask image as the input image at the recognition of moving objects using the HMM.

After this procedure, the coarse detected mask image is given as the input image to HMM for recognition of the object.

2.2 Recognition of Object

This section will focus on the method used for recognizing the object which is detected using Temporal Difference. This procedure is divided into two steps. First, we are going to extract feature vectors and calculate wavelet coefficients from the candidate region of the detected moving object by using Discrete Wavelet Transform. In the second step, we provide the feature vectors as input to Hidden Markov Model for object recognition. In this section, a short description of Discrete Wavelet Transform and Hidden Markov Model is provided.

A. Perform DWT for extracting feature vectors

When we are using discrete wavelet transform there are three basic steps involved to obtain a sequence of feature vectors from a detected mask image. First, a properly sized image is obtained by normalizing the detected mask image. This is important in the sense that it can accurately capture the geometry characters of the moving objects. In the second step, we obtain a sequence of sub-images for the given normalized image by capturing over the object image in a raster scan manner, with a square window of fixed size and a predefined overlap ratio. Thus, relevant information on the local geometry of the object to be encoded can be captured. Also the sequence of subsequent windows summarizes the structure of local object. Finally, we apply the two-dimensional Haar basis wavelet transformation technique to each sub-image. This technique can be used to decompose a given image into four sub-images [1].

Using the four sub-images obtained above, it is seen that the wavelet transformation technique can preserve not only feature vectors but also spatial ones. We can decompose an original image into four bands (LL, HL, LH, HH) by using the DWT technique [1,2,10]. These sub-bands exhibit different frequency characteristics through high-pass and low-pass filters. High-pass filters extract the high-frequency information, and the low-pass filter provides low-frequency information representing the highest energy of the image. The wavelet coefficients for the image are calculated and this is followed up by their sorting in decreasing order of magnitude. Subsequently, for the first M coefficients, we can retain those of higher magnitude, performing lossy image compression. Dimensionality of the observation vector is determined by the number of retained coefficients, and the number of sub-images is determined by the length of the vector. By applying this step to all the sub-images in the sequence, we finally obtain the feature vectors whose dimensionality is $M \times T$, where M is the number of wavelet coefficients retained and T is the number of sub-images gathered during the sample scanning operation and these are used with HMM. The HMM, is a good method for capturing the sequential nature of data, can successfully use the recognition of an object from the unrolled sequence of its wavelet coefficients.

B. HMM for recognizing the moving objects

By combining wavelet coefficients with the HMM, a powerful framework can be constructed for classification of the objects. The standard way for implementing HMM is $\lambda = (\pi, A, B)$ to recognize each class of moving objects. When a set of sequences of wavelet vectors $\mathbf{O} = \{\mathbf{O}_1, \mathbf{O}_2, \dots, \mathbf{O}_T\}$ is given, the training of the model is typically performed using the standard Baum-Welch re-estimation method, which determines the parameters $\lambda = (\pi, A, B)$ of the HMM that maximize the likelihood function $P(\mathbf{O} | \lambda)$. For each class of objects one model could be trained and if a new observation sequence $\mathbf{O}$ is given, the forward-backward procedure is used to compute the likelihood function $P(\mathbf{O} | \lambda)$ for each trained HMM $\lambda = (\pi, A, B)$. The classification of objects is performed using the standard maximum likelihood classification rule, which assigns an unknown item to the class whose model shows the highest likelihood [5,10].

This method may be considered as nearest neighbor classifier, with the proximity measure defined by the likelihood function [10]. For the object recognition problem the above classification scheme is appropriate. We look for the most similar view, to identify an object with a given aspect that is most likely to represent the near view of the same object. In our case after training one model for each aspect of the object, an unknown view is assigned to the class of the most similar view, in which the similarity is computed by using the likelihood of the HMM. We do not need to retrain the class model if a new view of the object is added which is a key advantage of the scheme.

III. CONCLUSION

This method proposes a new approach for detection of moving object using TDM which is relatively exact and robust to background noise and clutter. This system rapidly detects the coarse region of moving object and is also suitable for dynamic background because of adaptive threshold. DWT and HMM is then used to effectively classify the detected object such as humans, animals or vehicles. The Clausius entropy domain has an attractive property in that it can decompose image into background in stable conditions and foreground at unstable conditions. This is observed that combination of Clausius entropy, Temporal Difference and Hidden Markov Model can significantly improve the detection and recognition performance of moving objects in video sequence.

REFERENCES

- [1] Ales Prochdzka, Andrea GavlasovB and Karel Volka, "Wavelet Transform in Image Recognition", Prague Institute of Chemical Technology, Czech Republic 2005
- [2] Chih-Hsien Hsia, "Improved Directional Lifting-Based Discrete Wavelet Transform For Low Resolution Moving Object Detection", National Taiwan University of Science and Technology Taipei city, Taiwan, 2012
- [3] Fu-Yuan Hu, Yan-Ning Zhang, Lan Yao, "An Effective Detection Algorithm For Moving Object With Complex Background", Northwestern Poly-Technical University, China 2005
- [4] Henan Guo, Yanchun Liang, Zhezhou Yu, Zhen Liu, "Implementation and Analysis of Moving Objects Detection in Video Surveillance", China, 2010
- [5] Hung-Son Le and Haibo Li, "Face Identification System using Single Hidden Markov Model and Single Sample Image per Person", Digital Media Lab, Sweden 2004
- [6] J. H. Park, G. S. Lee, W. H. Cho, N. Toan, S. H. Kim, S. Y. Park, "Moving Object Detection based on Clausius Entropy", 10th IEEE International Conference on Computer and Information Technology, 2010.
- [7] W. Cho, S. Kim, G. Ahn, "Detection and Recognition of Moving Objects using the Temporal Difference Method and the Hidden Markov Model", Chonnam National University, Gwangju, South Korea, 2011.
- [8] W. Cho, S. Kim, G. Ahn, S. Park, "Detection and Tracking of Multiple Moving Objects in Video Sequence using Entropy Mask Method and Fast Level Set Method", Dept. Statistics, Chonnam National University, 2011.
- [9] Wei Shuigen, Chen Zhen, "An Improved Method of Motion Detection Based on Temporal Difference", Nanchang Hangkong University Nanchang, Jiangxi, China, 2009
- [10] Wu Meng-yu, Han Ninj, Luo Qin-juan, "A Smoke Detection Algorithm Based On Discrete Wavelet Transform and Correlation Analysis", Beijing Forestry University, China, 2012.