Tracking-based Visual Surveillance System

Sandip A. Deore[1], Manish M. Patil[1]

[1]*Department of E&TC, Gangamai College of Engineering, Nagaon, Dhule(MS), INDIA*

Abstract *One of the current demanding research problems in computer vision is visual surveillance in dynamic situations, particularly for humans and automobiles. It is a critical technology in the battle against terrorism, crime, public safety, and effective traffic management. The endeavor entails the creation of an effective visual surveillance system for use in complex contexts. Detecting moving objects from a video is critical for object categorization, target tracking, activity recognition, and behavior understanding in video surveillance. The first relevant phase of information is the detection of moving objects in video streams, and background subtraction is a typical method for foreground segmentation. We simulated background subtraction and Self-Organizing Background Subtraction (SOBS) approaches in this paper. Establishing a credible background updating model based on statistical data and using a dynamic optimization threshold method to produce a more complete moving object is the goal of Background Subtraction. In SOBS, a method for detecting moving objects is based on a neural background model that is automatically produced via a self-organizing mechanism.*

Key Words: **Background Subtraction, Motion Detection, Neuronal Mapping, Self -Organizing Background Substraction, Visual Surveillance**

1. INTRODUCTION

Understanding the motion of moving objects in a picture using video is a difficult scientific challenge as well as a fruitful domain with numerous potential applications. As a result, it has attracted the attention of a number of academic and commercial organizations. The study and development of moving object detection algorithms is our motivation. Moving Object detection is the first stage in subsequent video analysis. It is responsible for separating moving items from fixed background objects. This not only provides a center of attention for higher-level processing, but it also significantly reduces computing time. Background removal and statistical models are two commonly used strategies for object detection.

Object segmentation is a challenging and critical challenge that must be handled well for a robust visual surveillance system due to environmental factors such as lighting changes, shadows, and blowing tree branches in the wind. In our work, we used two methods for object recognition with background subtraction: a self-organizing approach to background subtraction and a traditional approach to background subtraction. Tracking is the next phase in video analysis, which is simply defined as the development of temporal connection between observed objects from frame to frame. This approach generates cohesive information on the objects in the observed area, such as trajectory, speed, and direction, and offers temporal identification of the segmented regions. The tracking step's output is commonly used for higher-level activity analysis.

2. LITERATURE OVERVIEW

Private corporations, governments, and public organizations utilize automatic visual surveillance to combat terrorism and crime, as well as to ensure public safety in airports, bus stops, train stations, city centers, and hospitals. It has also found use in traffic surveillance for effective transportation network management and road safety. Motion detection, object classification, personal identification, tracking, and activity recognition are all tasks that can be performed by a visual surveillance system. Detection of moving objects is the first critical stage in the process indicated above, and accurate segmentation of moving foreground objects from the background ensures object classification, personal identification, tracking, and activity analysis, making the subsequent steps more efficient. [6] divided motion detection into three categories: frame differencing, optical flow, and back ground subtraction. Frame differencing [7] detects regions belonging to moving objects such as humans and cars by pixel-wise differencing between two or three consecutive frames in a picture stream. Change is determined by the threshold function, which is dependent on the speed of object motion. When the object's speed fluctuates dramatically, it's difficult to maintain segmentation quality. Frame differencing is particularly adaptable to dynamic situations, however moving elements frequently produce holes.

The optical flow [5] technique detects moving parts in an image by tracking the flow vectors of moving objects over time. It is employed in applications requiring motion-based segmentation and tracking. Each pixel region's translation is defined by a dense field of displacement vectors. In the presence of camera motion, optical flow works best, however most methods of computation are computationally expensive and sensitive to noise. The computation of optical flow [25] is only utilized in moving areas. In order to track moving objects, first the moving edge in polar-log coordinate is retrieved, and then the gradient operator, in polar-log coordinate, is used to compute the optical flow directly in the moving area.

The most popular and widely used strategy for motion detection is background removal [6], [7], [8], and [11]. Essentially, the idea is to take the current image and subtract it from a reference background image that is updated in real-time. But the principle only works if stationary cameras are present. Only non-stationary or new items remain after the subtraction, which include the whole silhouette region of an object. For real-time systems, this approach is straightforward and computationally economical, but it is particularly sensitive to dynamic scene changes caused by lightning and other events. As a result, it is extremely reliant on a robust background maintenance model.

Background subtraction finds moving regions in a picture by calculating the difference between the current image and a reference background image acquired over time from a static background. Only non-stationary or new items remain after the subtraction, which include the whole silhouette region of an object. Backdrop subtraction [4], [5] solves the problem of automatically updating the background from an incoming video frame, and it should be able to solve the following issues:

- **Background motion:** Identify non-stationary background regions such as tree branches and leaves, flags flapping in the wind, and flowing water as part of the background.

- **Variations in illumination:** Gradually changing illumination should allow the background model to adjust over time.

- **Variations in illumination:** Gradually changing illumination should allow the background model to adjust over time.

- **Memory:** The background module should not consume a lot of computing power and memory.

- **Camouflage:** Moving items should be recognized even when their pixel properties are identical to those of the background.

- **Bootstrapping:** The background model should be able to sustain it even when there is no training background (no foreground object).

Authors in [10] present a novel adaptive background subtraction method for segmenting moving regions and locating human body positions. So far, most of the approaches proposed alter the permitted range of background picture fluctuations based on background image training samples. As a result, the detection sensitivity of pixels with large permitted ranges drops. Background modelling and moving object segmentation [12] are important approaches for video surveillance and other video processing applications. Most known approaches to modeling background and segmenting moving objects operate primarily at the pixel level in the spatial domain.

The majority of statistical background subtraction approaches in [13] are based on a temporal color/intensity distribution analysis. Learning statistics from a series of time frames, on the other hand, can be challenging, especially if no frames without moving objects are available, or if the memory is not sufficient to record the required series of frames. In general,[14] the following processes are included in the processing framework of visual surveillance in dynamic scenes: modelling of surroundings, motion detection, classification of moving objects, tracking, interpretation and description of actions, human identification, and fusion of data from various cameras. The detection of moving targets in infrared video sequences is proposed in this paper using a novel progressive estimation approach [15]. There is no restriction on camera movements or requirement for a stiff background in the technique described.

This study develops the detection and tracking [17] of several targets with a camera. This study proposes a mixture Gaussian background model based on object-level for identifying moving targets from video sequences after background subtraction.

Authors in [2] and [8] discussed a SOBS issue. First, in many computer vision applications, moving objects must be detected in video streams in order to extract relevant information. Apart from the obvious benefit of being able to break down streams of video into moving and background components, recognizing moving objects facilitates recognition, classification, and activity analysis, making those processes more efficient. An artificial neural network-based technique to self-organization that is commonly used in human image processing systems.

Method in [18] is an example. The difficulty of real-time object tracking entails extracting important information from complicated and unreliable visual data. In this study, we show how to develop an artificial neural network (ANN) for a real-time object tracking application using a complete methodology. The item being tracked for demonstration purposes is a specific aero plane. The proposed ANN, on the other hand, can be taught to track any other object of interest.

A new feature extraction strategy for feed forward neural networks is described in this [19] paper. According to this method, all of the necessary classification characteristics can be derived from the decision boundary, which is based on the boundary feature extraction algorithm. Feature extraction from decision boundaries can take advantage of the properties of neural networks to handle complex issues with arbitrary decision boundaries without assuming the underlying probability distribution functions.

Authors in [16] present a self-organizing technique based on artificial neural networks, which is widely used in human image processing systems and more broadly in cognitive research. The suggested model captures structural background variation over a prolonged period of time due to periodic-like motion when memory is limited. Our system is capable of handling situations with moving backdrops or lighting variations, as well as achieving reliable detection for various types of footage captured with stationary cameras.

3. PROPOSED FRAMEWORK

The most frequent approach of motion detection is background removal. It is a technology that detects motion using the difference between the current image and the backdrop image, and can often provide data such as object location. The initialization and updating of the backdrop image are the most important parts of this approach. The efficacy of both will have an impact on the accuracy of test results. The result is that this method employs a real-time initialization and updating of the backdrop in an efficient manner.

The original backdrop image might be obtained in a variety of ways. For instance, the first frame may be used as the background directly, or the average pixel brightness of the first several frames could be used as the background, or background models can be estimated using image sequences without moving objects, and so on.

The time average approach is the most widely used way for establishing an initial background among these strategies. However, this strategy is ineffective when dealing with a background image that has shadow issues. The method of calculating the median from a continuous multi-frame can effectively and easily solve this problem.

As a result, the median approach is used to initialize the background in this procedure. The following is an example of expres

$$Binit(x, y) = mediamfk\ (x, y)\ k = 1, 2, ...n \quad (1)$$

Where Binit is the initial background, n is the total number of frames selected.

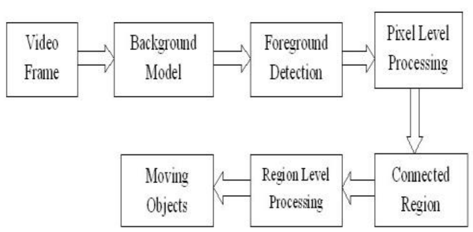

Fig. 1 Proposed Framework

Where Binit is the initial background, n is the total number of frames selected.

In order for the backdrop model to effectively react to light changes, the background must be updated in real time so that the moving item can be accurately extracted. The updating algorithm in this method is as follows: When a moving item is detected, the pixels that are determined to belong to the moving object keep their original background grey values and are not changed. Update the background model according to the following rules for pixels that are determined to constitute the background:

$$Bk+1 = \beta Bk\ (x, y) + (1 - \beta)Fk\ (x, y) \quad (2)$$

Where $\beta \epsilon (0, 1)$ is update coefficient, in this method β = 0.004. Fk (x, y) is the pixel gray value in the current frame. Bk (x, y) and Bk+1(x, y) are respectively the back- ground value of the current frame and the next frame.

The background image B(x, y) subtracted from background image B(x, y) from the current frame Fk (x, y). If the pixel difference is reaches the threshold T , then it the pixels appear in the moving object, else it is background pixels. The moving object can be detected after threshold operation. Its expression is as follows:

$$Dk\ (x, y) = 1 \qquad |F\,k(x, y) - Bk{-}1(x, y)| > T \quad (3)$$

Dk (x, y) is the binary image of differential results. T is gray-scale threshold that determines the object identification accuracy.

As the complexity of the background, the difference image obtained contains the motion region, in addition, also a large number of noise. Therefore, noise needs to be removed. In this method median filter with the 3X3 window and filters out some noise. The motion zone may comprise moving autos, flying birds, flowing clouds, swaying trees, and other non-body parts in addition to body parts.

Morphological methods are used for further processing. Firstly, corrosion operation is taken to effectively filter out non-human activity areas. In addition, the expansion operation can be used to filter out most of the non-body motion regions while pre-serving the human motion without causing injury. Some isolated spots of the image and some interference of small pieces are eliminated after expansion and corrosion operations, and we get a more accurate human motion region.

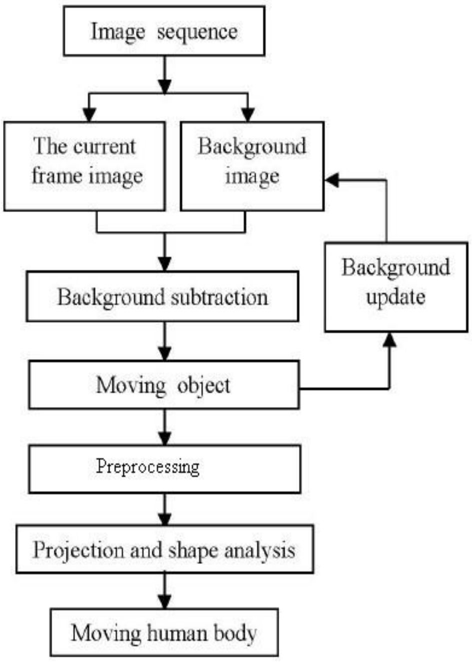

Fig. 2. Moving Object Extraction

Some precise edge areas will be obtained after preprocessing, but the region that corresponds to the moving human body cannot be established. We can see that when a moving object appears, it casts a shadow in some areas of the scene. The existence of shadow will have an impact on the accuracy of the moving object extraction. After examining the characteristics of motion detection, we combine the projection operator with the preceding approaches. Adopting the approach of combining vertical and horizontal projection to identify the height of the motion region based on the results of the methods above. This can eliminate the impact of the shadow to a certain degree.

The background model utilized in the SOBS technique is based on the idea of generating an image sequence neural background model by self-organizing learning picture sequence variations as pixel trajectories in time. The network functions as a competitive neural network with a winner-take-all mechanism, as well as a mechanism that affects the local synaptic plasticity of neurons, this enables learning to be spatially confined to the immediate vicinity of the most active neurons. As a result, the neural background model adapts efficiently to scene changes and can capture the most consistent element of the image sequence.

In the case of our background modeling application, we have at our disposal a fairly good means of initializing the weight vectors of the network: the first image of our sequence is indeed a good initial approximation of the background, and, therefore, for each pixel, the corresponding weight vectors are initialized with the pixel value.

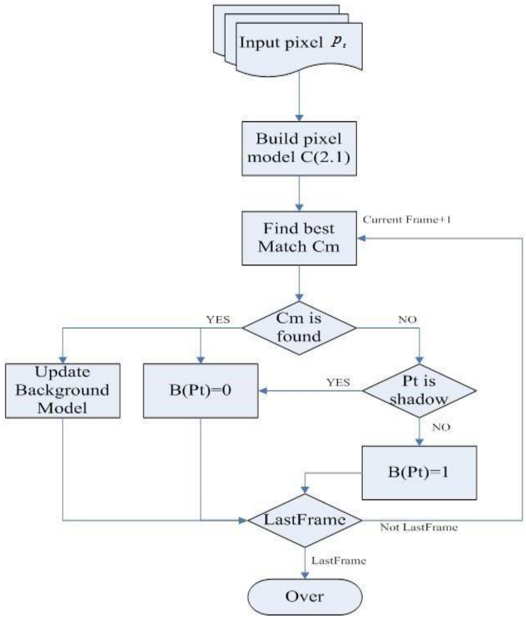

Fig. 3 SOBS Algorithm

Choose the HSV colour space to represent each weight vector, relying on the hue, saturation, and value attributes of each colour. This type of colour space enables us to characterise colours in a way that is similar to how humans perceive colours. Furthermore, the intensity of the light is explicit and distinct from chromaticity, making change detection insensitive to changes in illumination strength possible. Let (h, s, v) be the HSV components of the generic pixel of the first sequence frame, and let C = (C1, C2, ..Cn2) be the model for pixel (x, y) .

Following initialization, the network is supplied with temporally consecutive samples. Each arriving pixel point of the tth sequence frame is compared to the current pixel model to see if a weight vector that best matches it exists. If a best matching weight vector is found, it means that belongs to the background and it is used as the pixel encoding approximation, and the best matching weight vector, together with its neighborhood, is reinforced.

If no appropriate matching weight vector exists, determine whether or not it is in the shadow cast by some object. In the first case, it should still be considered background, but it should not be used to update the corresponding weight vectors in order to avoid the reinforcement of shadow information into the background model; in the second case, it is detected as belonging to a moving object and should not be used to update the corresponding weight vectors (foreground). The following algorithm can be used to sketch the background subtraction and update procedure mentioned above for each pixel.

To determine which weight vector gives the best match, several metrics for detecting changes in color imagery, such as those reported in and in references therein, could be adopted. The Euclidean distance of vectors in the HSV color hexcone that gives the distance between two pixels pi = (hi, si, vi) and pj = (hj , sj , vj) as

$$d(p_i, p_j) = \|(v_i s_i \cos(h_i), v_i s_i \sin(h_i, v_i) - (v_j s_j \cos(h_j), v_j s_j \sin(h_j, v_j)\| \quad (4)$$

The representation of HSV values as vectors in the HSV color hexcone used in such distance measure allows to avoid problems connected with the periodicity of hue (that represents an angle) and with the instability of hue for small values of saturation (hue is undefined for null saturation)The threshold allows to distinguish between foreground and background pixels, and is chosen as $\epsilon = \epsilon 1$, if $0 \leq t \leq K$

If a best matchCm is found for current samplept , the weight vectors in the neighborhood of Cm are updated according to selective weighted running average. In details, given the incoming pixel pt(x, y) at spatial position(x, y) and timet , if there exists a best matchCm in the model C of pt , and is present in the background model at position , then weight vectors in the neighborhood of are updated according to

$$A_t(i, j) = (1 - \alpha_{i,j}(t))A_{t-1}(i, j) + \alpha_{i,j}(t))p_t(x, y) \quad (5)$$

If the best match Cm is not found, the background model remains unchanged. This selectivity allows to adapt the background model to scene modifications.

Fig. 4. Current Image

Fig. 5. Difference Image

Fig. 6. Fixed Threshold Result

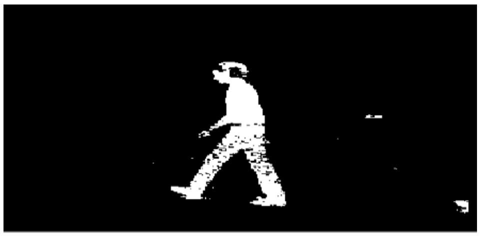

Fig. 7 Dynamic Threshold Result

Fig. 8 SOBS Output

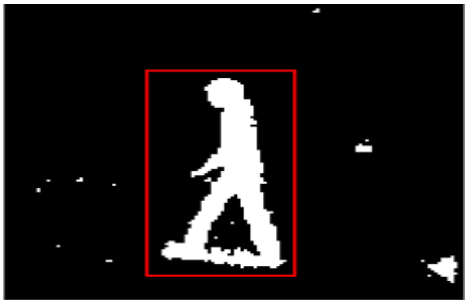

Fig. 9 Reprocessing Result

4. RESULT AND DISCUSSION

We reviewed the implementation aspect of background subtraction and the self-organizing background subtraction method utilized for moving object recognition in the previous chapter. Background subtraction is a technique for separating objects of interest in a scene from the background. The Self Organizing Backdrop Subtraction method of identifying moving objects is based on a background model that is automatically generated by a self-organizing process without any prior knowledge of the patterns involved. We will compare these two approaches using various characteristics such as MSE, PSNR, Processing Speed, and Entropy in this paper.

Results		
Name	Background Subtraction	SOBS
MSE	661.3294	1174.3
PSNR	19.9267	17.0685

Table: 1 Results

5. CONCLUSIONS

Background subtraction and Self-Organizing Background Subtraction algorithms were used to detect moving objects. While background subtraction is a more reliable method than SOBS. To compare results between a ground truth image and the output we're getting. The MSE, PSNR, and processing time are used to compare the results. When compared to SOBS, the processing time required for background subtraction is reduced.

We create a trustworthy background model, apply dynamic threshold approach to detect moving objects, and update the background in real time using background subtraction method to address deficiencies in the traditional way of object identification.

REFERENCES

[1] Lijing Zhang and Yingli Liang,"Motion human detection based on background Subtraction",Second International Workshop on Education Technology and Com- puter Science, pp 284-287, Year: 2010.

[2] Lucia Maddalena and Alfredo Petrosino "A Self Organizing Approach to Back- ground Subtraction for Visual Surveillance Applications", IEEE Transactions, Vol.17, Issue.7, Year: 2008.

[3] Massimo Piccardi "Background subtraction techniques: a review",IEEE Interna- tional Conference on Systems, Man and Cybernetics, pp. 3099-3104, Year:2004.

[4] O. Barnich and M. Van Droogenbroeck. ViBe:"A Universal Background Subtrac- tion Algorithm for Video Sequences", IEEE Transactions on Image Processing, Vol. 20, Issue.6, pp. 1709 -1724, Year: 2011.

[5] Ahmed Elgammal, David Harwood, and Larry Davis." Non-parametric model for background subtraction", In Frame Rate Workshop, IEEE, pp.751-767, Year: 2000.

[6] Weiming Hu, Tieniu Tan, Liang Wang, and S. Maybank. " A Survey on Visual Surveillance of Object Motion and Behaviors", IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), Vol.34, Issue.3, pp. 334-352, Year: 2004.

[7] Mingxin Jiang, Min Li and Hongyu Wang, "A Robust Algorithm of Object Tracking Based on Moving Object Detection", In International Conference on Intelligent Control and Information Processing, pp. 619-622, Year: 2010.

[8] Lucia Maddalena and Alfredo Petrosino "The SOBS algorithm: what are the limits ?" ,pp. 21-26, IEEE , Year:2012.

[9] M. Dimitrijevic, "Human body pose detection using Bayesian spatiotemporal tem- plates", International Conference on Intelligent and Advanced Systems, pp.764-769, Year: 2008.

[10] Tao Jianguo, Yu Changhong, "Real-Time Detection and Tracking of Moving Object", Intelligent Information Technology Application, UTA '08, Second Inter- national Symposium, Vol. 2,pp.860-863, Year:2008.

[11] Niu Lianqiang and Nan Jiang, "A moving objects detection algorithm based on improved background subtraction", Intelligent Systems Design and Applications, ISDA '08. Eighth International Conference, Vol.3, pp 604 - 607, Year: 2008

[12] Wang Weiqiang, Yang Jie and Gao Wen, "Modeling Background and Segment- ing Moving Objects from Compressed Video", Circuits and Systems for Video Technology, IEEE Transactions , Vol. 18, Issue 5, pp. 670-681,Year:2008

[13] M. Mignotte and I Konrad, "Statistical Background Subtraction Using Spatial Cues", Circuits and Systems for Video Technology, IEEE Transactions, Vol. 17, Issue 12, pp. 1758 -1763,Year:2007.

[14] Weiming, H., Tieniu, T., Liang, W., and S, M. "A survey on visual surveillance of object motion and behaviors", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, pp. 334-352.Year: 2004.

[15] Lipton, H. Fujiyoshi, and R. Patil, "Moving Target Detection and Classification from Real-Time Video", In Proceedings of IEEE Workshop on Application of Computer Vision,Year: 1998.

[16] L. Maddalena and A. Petrosino, "A self-organizing approach to detection of moving patterns for real time applications", In Proc.2nd Int. Symp. Brain, Vision ,and Artifcial Intelligence, pp.181190, Year:2007.

[17] S. C .Cheungand C. Kamath, "Robust techniques for background subtraction in urban traffic video", in Proc. EI-VCIP, pp.881892, Year: 2004.

[18] Javed Ahmed, M. N. Jafri, J. Ahmad, and Muhammad I. Khan "Design and Implementation of a Neural Network for Real Time Object Tracking", World Academy of Science Vol. 6, Engineering and Technology, Year: 2007.

[19] C. Lee and D. A. Landgrebe, “Decision boundary feature extraction for neural networks”, IEEE Transactions on Neural Networks, Vol.8, No.1, pp.75-83, Year:1997.

[20] David Demirdjian, T. Ko, and Trevor Darrell. “Constraining human body track- ing”, In Proc. IEEE Int. Conf. on Computer Vision, Vol. 2, pp. 1071-1078, Year: 2003.

[21] Zhong, Anil K. Jain, M.-P. Dubuisson-Jolly, “Object Tracking Using Deformable Templates”, IEEE transactions on pattern analysis and machine intelligence, Vol. 22, no. 5, pp. 544-549, Year: 2000.

[22] Xuechao. Y, and Wenping L, “Moving objects detection technology in video sequences”, Computer Applications and Software, pp. 215-217, Year 2008.

[23] Du-Ming Tsai and Shia-Chih Lai, “Independent Component Analysis Based Background Subtraction for Indoor Surveillance”, Image Processing, IEEE Transactions, Vol. 18, Issue 1, pp.158 - 170 Year 2009.

[24] C. Stauffer and W. E. L. Grimson, “Adaptive background mixture models foe realtime tracking”, in IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, Vol. 2,pp. 246252, Year:1999.

[25] P. Kaew, T. K. Pong, and R. Bowden, “An improved adaptive background mixture model for real-time tracking with shadow detection”, in Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems, Year: 2001.

[26] Hai-Yan Zhang, “Multiple moving objects detection and tracking based on optical flow in polar-log images”, Machine Learning and Cybernetics (ICMLC), 2010 International Conference, Vol. 3, pp. 1577 - 1582, Year: 2010.

[27] A. Talukder, L. Matthies, “Real-time detection of moving objects from moving vehicles using dense stereo and optical fl w ”, Intelligent Robots and Systems, pp. 3718 - 3725 Vol.4, Year: 2004.

[28] J. L. Barron, D .J. Fleet, and S. S. Beauchemin, “Performance of optical flow tech- niques”, Int. J. Comput.Vis., Vol.12, No.1,pp.4277, Year: 1994.