DOI: 10.21917/ijivp.2017.0204

OBJECT TRACKING WITH ROTATION-INVARIANT LARGEST DIFFERENCE INDEXED LOCAL TERNARY PATTERN

J. Shajeena[1] and K. Ramar[2]
[1]*Department of Computer Science and Engineering, James College of Engineering and Technology, India*
[2]*Einstein College of Engineering, India*

Abstract
This paper presents an ideal method for object tracking directly in the compressed domain in video sequences. An enhanced rotation-invariant image operator called Largest Difference Indexed Local Ternary Pattern (LDILTP) has been proposed. The Local Ternary Pattern which worked very well in texture classification and face recognition is now extended for rotation invariant object tracking. Histogramming the LTP code makes the descriptor resistant to translation. The histogram intersection is used to find the similarity measure. This method is robust to noise and retain contrast details. The proposed scheme has been verified on various datasets and shows a commendable performance.

Keywords:
LTP, Motion Vector, Rotation-Invariant, Histogram, Object Tracking, DCT

1. INTRODUCTION

Object tracking commends a vital topic in computer vision and pattern recognition. Its main objective is to analyse videos and to detect certain unusual activities [3].

Object tracking is active in video surveillance systems specifically to track person and vehicles. Venkatesh Babu and Anamitra Makur have worked with video surveillance using object based video compression system using foreground motion compensation. This method segments the moving objects independently from the video and codes them with respect to the previously reconstructed frame [4].

During the byegone years the Local Binary Patterns (LBPs) has peaked its popularity because of its high discriminative nature, its key advantages and its invariance to monotonic gray level changes and computational efficiency.

Initially the texture description used the LBP operator. Zhenhua Guo et all proposed a complete modelling of LBP operator for texture classification [5][6]. LBP has been popularly used for face recognition [7] and feature extraction [8]. A popular robust version of LBP is used in [9] shows a superb performance on human face recognition, which address the challenge of improving the robustness to image noise. LBP based facial image analysis has been the topper and successful application during bye gone years.

Furthermore some methods retain the contrast information of image patterns. They contain both edge and texture information which is desirable for object recognition. In [10] authors use LBP for object recognition.

The Local Binary Pattern operator, used in many domains such as texture classification and face recognition has been applied to object tracking [11][12]. Here the Local Binary Pattern (LBP) histogram pattern of each image in the sequence and the reference pattern is constructed.

Anbarasa Pandian et al. [13] have proposed a feature extraction technique for MRI brain tumor image in two different steps namely, feature extraction process and classification. The curvelet transform, contourlet transform and local ternary pattern techniques are used for texture feature extraction. Supervised learning algorithm like Deep Neural Network (DNN) is used for classification.

The LTP operator has shown its superiority in face recognition [14][15]. It is used in the presence of noise and deal with the problem of generalization.

Some methods such as [16][17] have also addressed the problem of rotation invariance using LBP features. Rotations of an input image cause the LBP patterns to translate into a different location and rotate about their origin.

In this paper, we propose an object tracking technique which uses LDILTP operator, where bit patterns are considered from the largest difference. Finally for classifying histogram intersection measure is used.

The rest of the paper is organized as follows: Section 2 explains the proposed object tracking method that includes various techniques such as DCT compression, Local Ternary Pattern and estimation of motion vector. The performance evaluation metrics are detailed in section 3, experimental results are discussed in section 4. Section 5 concludes the paper.

2. PROPOSED METHOD

The proposed object tracking technique is given in a step by step procedure as:

1. Capture the frames from the video clip.
2. Apply low pass filter on the captured frame.
3. Transform the captured RGB frame to YCbCr color space.
4. Divide the Y plane of the YCbCr frame into 8×8 blocks.
5. Find DCT coefficients for each block.
6. Extract DC coefficients which results in $m/8 \times n/8$ matrix F_i for the ith frame.
7. Find the frame difference $\delta_i = |F_i+1 - F_i|$.
8. Detect the area of movement (ROI) by checking if $\delta_i > t$.
9. Extract the features by taking LDILTP for ROI of F_i.
10. Apply morphological operations Dilation and Erosion to fill the gap in ROI.

11. Label ROI data and find the largest component which is set as object area.
12. Track the object area in ROI using Objet Track Feature Frames.
13. Evaluate the performance of the system.

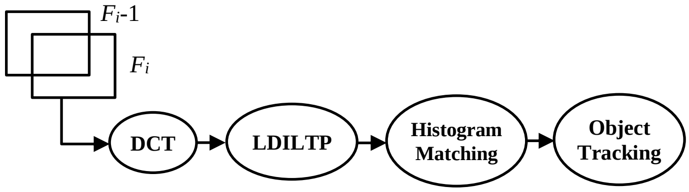


Fig.1. The overview of the proposed object tracking technique

The Largest Difference Indexed Local Ternary Pattern, temporal tracking model is the process of selecting the motion vector, to transform into Discrete cosine coefficients and track each object roughly in real time.

The methodology for object tracking can be divided into four broad steps, namely, DCT compression, Feature extraction using LDILTP, Morphological processing using Fill Gap and Object Tracking. YCbCr is one of the popular color space in computing, also to have a more efficient representation of images, the RGB image is converted into YCbCr color space and the Y component is taken for further processing.

2.1 DCT COMPRESSION

The JPEG compression provides good visual quality in photo-real images. JPEG relies on quantization of the Discrete Cosine transform applied to 8×8 pixel blocks of an image.

Our proposed work uses Discrete Cosine Transform (DCT) on the Y plane in object tracking framework. A detailed explanation on DCT is given in our previous work [1][2]. The outcome of this step is an $m/8 \times n/8$ matrix of DC coefficients F_i for the ith frame. The area of movement is computed by finding the frame difference between the current frame and previous frame, and then checking it with threshold [1].

2.2 LOCAL TERNARY PATTERN (LTP)

When objects to be tracked from noisy, cluttered backgrounds, LTP operator is used. This operator is derived from Local Binary pattern (LBP). This operator works with the eight neighbors of a pixel, the center pixel as a threshold, and uses the resulting binary-valued image as a local image descriptor. Compared to Local Binary pattern (LBP), Local Ternary Pattern has 2 thresholds which creates three different states.

The LTP code at location (x,y) is computed as,

$$LTP_{x,y} = \sum_{b=0}^{B-1} S\left(P_b - P_c, T\right) 3^b \tag{1}$$

where, P_c is the pixel value at (x,y) P_b is the N_8 neighboring pixel values around P_c and B is the total number of neighboring pixels. 3^b bin block histogram is computed. $S(Z,T)$ is the threshold function where, T is a pre-defined threshold. Here we have empirically selected the threshold value as 0.1.

$$S\left(Z,T\right) = \begin{cases} 1, & z \geq T \\ 0, & -T < z < T \\ -1, & z \leq -T \end{cases} \tag{2}$$

The range of LTP histogram is very high. For B = 8, the histograms has 3^8(6561) bins. So for simplicity the LTP code is split into two separate channels of LBP descriptors, positive LBP code and negative LBP code. The positive LTP code and the negative code are obtained by:

$$S_p\left(Z,T\right) = \begin{cases} 1 & \text{if } z \geq T \\ 0 & \text{if } z < T \end{cases} \tag{3}$$

$$S_n\left(Z,T\right) = \begin{cases} 1 & \text{if } z \leq -T \\ 0 & \text{if } z > -T \end{cases} \tag{4}$$

for which separate histograms and similarity metrics are computed.

2.3 FEATURE EXTRACTION WITH LARGE DIFFERENCE INDEXED LOCAL TERNARY PATTERN (LDILTP)

The LTP operator is less sensitive to noise, as it encodes the small pixel difference into a separate state. But the dimensionality of LTP histogram is quite large. Large information may be lost while splitting into positive LBP code and negative code. Also the two histograms might hold redundant information.

Also while using the threshold function, the states, 0 and 1 might change from 0 to 1 or from 1 to 0 due to noise. Moreover when there is a rotation of image, the circular neighborhood on pixels also rotates, which produces erroneous results. To overcome all these limitations we have proposed LDILTP, which is rotation invariant.

The rotation invariant LDILTP descriptor, labels the pixels of YCbCr image by thresholding the pixels in a 3×3 neighborhood of each pixel with its central pixel value. The multiplier matrix is a matrix consists of 3^b. Now the multiplier matrix is shifted so that the largest difference value aligns with the largest multiplier. Multiply and sum to find LDILTP value.

11	12	13
14	15	16
17	18	15

(a)

4	3	2
1		-1
-2	-3	0

(b)

1	1	1
1		-1
-1	-1	0

(c)

27	9	3
81		1
243	729	2487

(d)

2487	729	243
1		81
3	9	27

(e)

	3367	

(f)

Fig.2. LDILTP Computation. (a). The original image window (b). the result after applying Eq.(1) (c). Binarized result after thresholding (d). Multiplier matrix (e). Shifted matrix (f). LDILTP value

Algorithm of Feature Extraction using LDILTP

Input: YCbCr frame

Initialization: ltpimg
Multiplier matrix

For each $X \leftarrow 1$ to total size of image
For each $Y \leftarrow 1$ to total size of image
Partition the region into 3×3 disjoint blocks
Compute the difference by subtracting the center pixel value.
Thresholding by applying the thresholding function, S
Shift the multiplier so that the largest difference value align with largest multiplier
Compute the dot product to find LDILTP value
Endfor
Endfor

Output: LDILTP Feature vector

The LDILTP is rotation invariant. For illustration, calculation of feature vector with LDILTP is shown with sample values in Fig.3. LDILTP1 shows pixels values in a 3×3 neighborhood and LDILTP2 shows a rotated version of the same pattern. The observation illustrates that the LDILTP is rotation invariant.

2.4 FILL GAP

When the object area produced by simple thresholding has several components with imperfections, then these regions are distorted by noise and texture. Morphological image processing operations comes in handy to regenerate the structure of the image.

The dual operations dilation and erosion, gradually enlarges the boundaries of regions of foreground pixels.

$$g = f \ominus s \tag{5}$$

$$g = g \oplus s \tag{6}$$

where, s is the structuring element, g is the output binary image and f is the input binary image.

LDILTP$_1$

1	2	3
4	5	5
2	7	6

Largest →

4	3	2
1	0	0
3	-2	-1

1(0)	1(1)	1(2)
1(7)	0	0(3)
1(6)	-1(5)	-1(4)

LDILTP$_2$

3	5	6
2	5	7
1	4	2

Largest →

2	0	-1
3	0	-2
4	1	3

1(2)	0(3)	-1(4)
1(1)	0	-1(5)
1(0)	1(7)	1(6)

LDILTP$_1$ = $1*3^0 + 1*3^1 + 1*3^2 + 0*3^3 + (-1)*3^4 + (-1)*3^5 + 1*3^6 + 1*3^7$
= 1+3+9+0-81-243+729+2487
= 2905

LDILTP$_2$ = $1*3^0 + 1*3^1 + 1*3^2 + 0*3^3 + (-1)*3^4 + (-1)*3^5 + 1*3^6 + 1*3^7$
= 1+3+9+0-81-243+729+2487 = 2905

Fig.3. LDILTP Computation for rotation invariance

2.5 ESTIMATION OF MOTION VECTOR

Motion estimation is an important step in the temporal difference learning technique used in MPEG video. The motion vector helps to fetch the rough estimation of the object region. Thus this associated information helps to search an acceptable match. Thus in compressed domain the moving object can be located in a limited region. Also the prior information about the motion vector in the compressed domain is a kind of unstructured information which can be obtained during the process of video uncompressing without any additional computation.

Fig.4. The Binary image after fill gap

From the detected ROI, each blocks are labeled and the second largest fragment is selected and searched in the next frame.

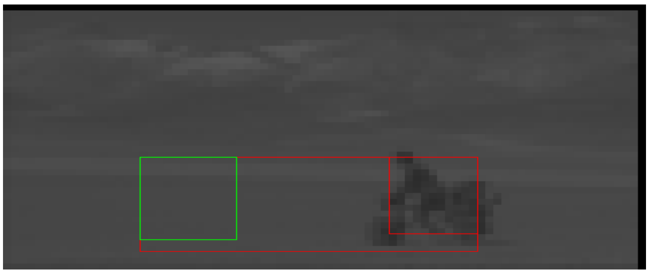


Fig.5. Object Area - first frame

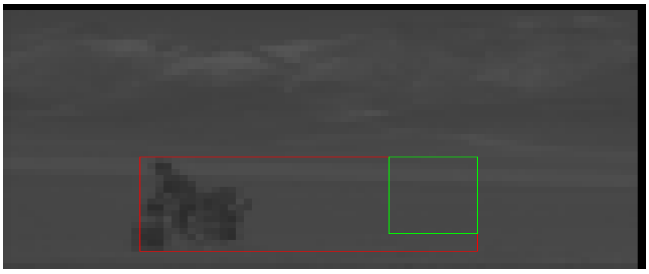


Fig.6. Object Area - second frame

Algorithm for Object Tracking using Object Track Feature Frames

Input: Coordinates of ROI i_x, i_y, f_x, f_y
Coordinates of object area in frame1 i_{x1}, i_{y1}, f_{x1}, f_{y1}
Calculate the histogram of object area

objHist := histogram of object area
Initialize current similarity score to a minimum value
For each $X \leftarrow i_x$ to f_x-objwidth
For each $Y \leftarrow i_y$ to f_y-objwidth
tmpBlk := frame2(x:x+objwidth-1,y:y+objheight-1)
tempHist := Histogram of tmpBlk
Evaluate new similarity score by histogram intersection as follows
objHist ∩ temphist := ∑ min(objHist, tempHist)
Largest similarity score is the best matching block
If new similarity score is higher than current similarity score
Set new similarity score as current similarity score
x of best matching block :=X
y of best matching block :=Y
Endif
Endfor
Endfor
Output: Motion vector x-i_x, y-i_y

2.6 SIMILARITY MEASURE

The LTP method for object tracking divides the image into a regular grid of cells and histograms are generated. The histogram intersection measure is a similarity metric that computes the local structure and geometric variations of such images.

To track the moving object, a sample(S) and a model(M), the difference between the feature vector is measured.

$$H\left(S,M = \sum_{b=1}^{B}\min\left(S_b,M_b\right)\right) \tag{7}$$

For a block after finding LTP, find the histogram match in the next frame to check for any histogram matches if any. In intersection the block with the maximum similarity is considered as match block.

3. PERFORMANCE METRICS

The experimental results show that the proposed method performs well. The algorithms were implemented in IDL6.3. The datasets for analysis were included from different tracking environments, about 30 video sequences were tested. The videos are manually recorded by fixing the camera with the object alone in motion and images are captured frame by frame using the standard tools. All of our experiments were performed on a standard PC (Intel i3 2.0 GHz CPU with 4 GB RAM) and with the various the resolution test videos.

In order to illustrate the performance of the proposed method, sensitivity, specificity, accuracy and error percentage were adopted to calculate the accuracy [3].

Sensitivity represents the trustfulness of a detection and is defined as the number of TP divided by the total number of labeled pixels, i.e., the sum of TP and FP. Specificity corresponds to the true negative rate and is defined as the number of TN divided by the total number of ground truth labels, i.e., the sum of TN and FP, where TP is the total number of true positives, FP is the total number of false positives, TN is the total number of true negatives and FN is the total number of false negative pixels.

		Manually Detected Object	
		Pixels Present	Pixels Absent
Detected by Proposed Method	Pixels Present	**TP**	**FP**
	Pixels Absent	**FN**	**TN**

Fig.7. Performance metrics

$$Sensitivity = \frac{\text{TP}}{\text{TP+FN}} \tag{8}$$

$$Specificity = \frac{\text{TN}}{\text{TN+FP}} \tag{9}$$

$$Accuracy = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} *100 \tag{10}$$

$$Error\ rate = 100 - A \tag{11}$$

Computation time, error percentage, accuracy, sensitivity and specificity are tabulated for different samples.

The average computation time of the proposed method is 35 seconds. The performance of the implemented algorithm is evaluated systematically using statistics definitions like true positives and true negatives. The Fig.7 gives the pictorial representation of the usage of statistical measures in the proposed method. Hence the ground truth of the tested video sequence was segmented manually and the metrics were evaluated.

The true positive is the number of pixels in the detected object area which is matched with the pixels in the manually segmented ground truth region. The true negative is the number of pixels not present in both the manually segmented image and the detected object. False positive is the number of pixels not present in manually segmented image but in detected object area. False negative is the number of pixels present in manually segments image and not in detected object area.

Table.1. Quantitative evaluation of test results

Sample	Computation Time	Error %	Accuracy	Sensitivity	Specificity
Sample1	27.5	1.743	98.256	0.8958	0.9835
Sample2	42.56	2.690	97.309	0.8045	0.9790
Sample3	43.87	0.583	99.41	0.8371	0.9955

Our proposed method is compared the tracking performance with one of our previous work [1] and the methods of Tamije Selvy et al. [18] and Sayed et al. [19] in terms of Accuracy and Sensitivity. The experimental results are listed in Table.2 and Fig.8. The results show that the proposed method has better performance in terms of accuracy amongst the other methods.

Table.2. Comparison with the state of the art methods

Method	Accuracy %	Sensitivity %
SVM	92	94.17
ELM	92.8	94.77
Block Based Tracking	96.63	88.96
Curvelet Transform	97.5	52
Contourlet Transform	97.5	52
LDILTP	98.32	84.58

4. EXPERIMENTAL RESULTS

The experimental results show that the proposed method can efficiently track moving objects in indoor and noisy outdoor environment with better accuracy. To make the proposed methodology more clear, some intermediary results for tracking the moving object are shown in Fig.9.

The Fig.9 details the experimental results of a sample case in an outdoor environment. From the video sequence the original video frames are captured, the first frame and second frame from the video sequence is displayed in Fig.9(a), Fig.9(b). The converted Y plane from the YCbCr color space is shown in Fig.9(c). To bring the captured frame to compressive domain, DCT of the Y plane is taken which provides a DC coefficient image Fig.9(d). The binary image of the detected ROI is shown in Fig.9€ which is the area of movement, computed by finding the frame difference between the current frame and previous frame, and then checking it with threshold. Manually segmented image in second frame is shown in Fig.9(f). The Fig.9(g), Fig.9(h) shows the object location in first frame and second frame. The tracked objects in first frame and second frame are labeled with arrows.

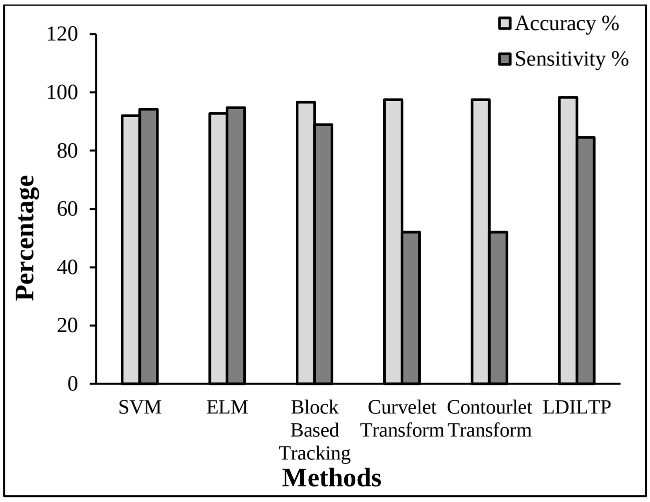
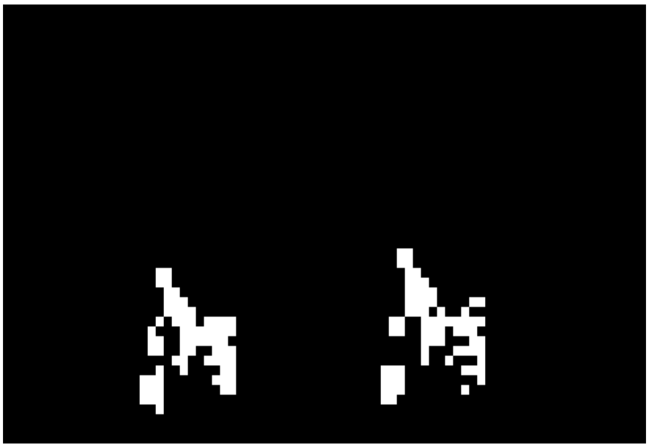


Fig.8. Performance comparison chart

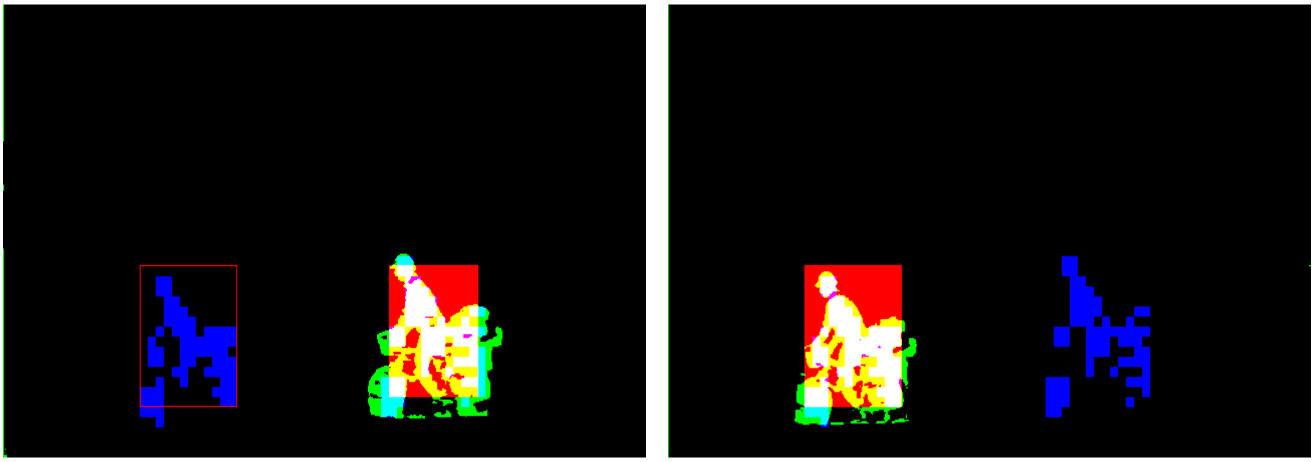
(a)

(b)

(c)



(d)



(e) (f)

Fig.9. Tracking a moving object - (a). Original frame1 (b). YCbCr image (c). DC coefficient image (d). Binary image of detected ROI (e). Tracked object first frame (f). Tracked object second frame

5. CONCLUSION

We have proposed an object tracking system in the compressed domain, using a rotation-invariant image operator called Large Difference Indexed Local Ternary Pattern (LDILTP) Tracking is based on temporal matching and using a histogram, the object movement is identified. The problem of rotation in the scenes is solved by indexing the matrix from the largest difference and shifting the multiplier.

Experimental results have shown that this approach obtains more accurate tracking results and tracks objects of various sizes and with large variations in illuminations even in outdoor surveillance scenarios.

Even though significant progress has been achieved for object tracking, several issues such as occlusion handling and tracking multiple objects may further be addressed.

REFERENCES

[1] J. Shajeena and K. Ramar, "Block-Based Tracking with Two Way Search", *ICTACT Journal on Image and Video Processing*, Vol. 5, No. 2, pp. 937-943, 2014.

[2] J. Shajeena and K. Ramar, "A Novel Way of Tracking Moving Objects in Video Scenes", *Proceedings of International Conference on Emerging Trends in Electrical and Computer Technology*, pp. 805-810, 2011.

[3] J. Shajeena and K. Ramar, "Object Tracking in Videos: Methodologies and Measures", *International Journal of Applied Engineering Research*, Vol. 10, No. 70, pp. 103-110, 2015.

[4] R. Venkatesh Babu and Anamitra Makur, "Object-based Surveillance Video Compression using Foreground Motion Compensation", Proceedings of 9th *International Conference on Control, Automation, Robotics and Vision*, 2006.

[5] Zhenhua Guo, Lei Zhang and David Zhang, "A Completed Modeling of Local Binary Pattern Operator for Texture Classification", IEEE Transactions on *Image Processing*, Vol. 19, No. 6, pp. 1657-1663, 2010.

[6] Taha H. Rassem and Bee Ee Khoo, "Completed Local Ternary Pattern for Rotation Invariant Texture Classification", *The Scientific World Journal*, Vol. 2014, pp. 1-10, 2014.

[7] Anuradha, Koneru and Manoj Kumar Tyagi, "A Novel Method of Face Recognition using Lbp, Ltp and Gabor Features", *International Journal of Scientific & Technology Research*, Vol. 1, No. 5, pp. 31-35, 2012.

[8] Anaghav Malkapurkar, Rupali Patil and Sachin Murarka, "A New Technique for LBP Method to Improve Face Recognition", *International Journal of Emerging Technology and Advanced Engineering*, Vol. 1, No. 1, pp. 67-71, 2011.

[9] A.A. Mohamed and R.V. Yampolskiy, "Adaptive Extended Local Ternary Pattern (AELTP) for Recognizing Avatar Faces", *Proceedings of International Conference on Machine Learning and Applications*, Vol. 1, pp. 57-62, 2012.

[10] Amit Satpathy, Xudong Jiang and How-Lung Eng, "LBP-based Edge-Texture Features for Object Recognition", *IEEE Transactions on Image Processing*, Vol. 23, No. 5, pp. 1953-1964, 2014.

[11] H. Rami, M. Hamri and Lh. Masmoudi, "Objects Tracking in Images Sequence using Local Binary Pattern (LBP)", *International Journal of Computer Applications*, Vol. 63, No. 20, pp. 19-23, 2013.

[12] Sk. Md. Masudul Ahsan, J.K. Tan, H. Kim and S. Ishikawa, "Spatiotemporal LBP and Shape Feature for Human Activity Representation and Recognition", *International Journal of Innovative Computing, Information and Control*, Vol. 12, No. 1, pp. 1-13, 2016.

[13] A. Anbarasa Pandian, and R. Balasubramanian, "Performance Analysis of Texture Image Retrieval for Curvelet, Contourlet Transform and Local Ternary Pattern Using MRI Brain Tumor Image", *International Journal in Foundations of Computer Science & Technology*, Vol. 5, No. 6, pp. 33-46, 2015.

[14] Abdallah A. Mohamed, and Roman V. Yampolskiy, "Adaptive Extended Local Ternary Pattern (AELTP) for Recognizing Avatar Faces", *Proceedings of 11th International Conference on Machine Learning and Applications*, Vol. 1, pp. 57-62, 2012.

[15] Jianfeng Ren, Xudong Jiang and Junsong Yuan, "Relaxed Local Ternary Pattern for Face Recognition", *Proceedings of 20th IEEE International Conference on Image Processing*, pp. 3680-3684, 2013.

[16] Zhenhua Guo, Lei Zhang and David Zhang, "Rotation Invariant Texture Classification using LBP Variance (LBPV) with Global Matching", *Pattern Recognition*, Vol. 43, No. 3, pp. 706-719, 2010.

[17] Guoying Zhao, Timo Ahonen, Jiří Matas and Matti Pietikäinen, "Rotation-Invariant Image and Video Description with Local Binary Pattern Features", *IEEE Transactions on Image Processing*, Vol. 21, No. 4, pp. 1465-1477, 2012.

[18] P. Tamije Selvy, T. Renuka Devi, R. Siva Keerthini and S. Umamaheswari, "A Proficient Extreme Learning Machine Approach for Tracking and Estimating Human Poses", *International Journal of Engineering and Computer Science*, Vol. 4, No. 3, pp. 10908-10913, 2015.

[19] Sayed Hossein Khatoonabadi, and Ivan V. Bajić, "Video Object Tracking in the Compressed Domain using Spatio-Temporal Markov Random Fields", *IEEE Transactions on Image Processing*, Vol. 22, No. 1, pp. 300-313, 2013.