International Journal of Engineering Research and Development
e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 12, Issue 5 (May 2016), PP.63-69

Content-Based Video Indexing and Retrieval Using Directional Temporal Local Binary Patterns (DTLBP)

M.Ravinder[1], Dr.T.Venu Gopal[2]
[1]*Research Scholar, JNTUK, Kakinada, Andhra Pradesh, India.*
[2]*Associate Professor, JNTUHCES, Sultanpur, Medak, Telangana, India.*

Abstract:- In this paper, we present a novel algorithm meant for content-based video indexing and retrieval, using directional temporal local binary patterns (DTLBP). DTLBP extracts motion information in 0^o, 45^o, 90^o, and 135^o directions. We have applied our proposed method on a data set of three hundred and thirty five videos (among which seventy two videos are of boat type; eighty videos are of car type; one hundred and forty eight videos are of airplane type, and thirty five videos are of war tank type) which are collected from TRECVID 2005, BBC and GOOGLE. To evaluate the performance of our algorithm, we have compared it with the existing volume local binary patterns (VLBP) method. Our method has shown reasonably good results compare to VLBP method.

Keywords:- Directional, temporal, texture, local binary patterns, volume local binary patterns.

I. INTRODUCTION

Video indexing and retrieval framework is required for annotating, storing, and retrieving videos by the end user. A video consists of rich content with raw data having minute predefined structure [1]. Due to the small size of video databases in olden days, videos have been annotated manually using keywords. However, in the present scenario, there is a rapid increase in size of video databases. Indexing manually, a huge video database is a trivial task. To overcome from this problem, we require a framework useful for indexing videos automatically based on the content of the videos. In the past literature, there have been a number of frameworks proposed for content-based video indexing and retrieval (CBVIR).

Temporal texture features are useful in finding motion information present in a video [1]. Temporal texture information represents the stationary features of the image sequence of a video [2]. Polana et al. in their paper [4] classified dynamic textures, activities, and events based on motion features. Chetverikov et al. in their survey paper [5] classified the temporal texture feature extraction methods as five types, and claimed that optical flow based methods were popular and successful, because of their efficiency in temporal texture feature extraction. Readers are redirected to [3-4, 6-13] for an extensive study on optical flow based methods. Fazekas et al. in [14] presented a comparison between normal flow characteristic features, and complete flow characteristic features, used for temporal texture classification. Saisan et al. in their paper [15] introduced a temporal texture recognition model, useful to identify fifty different dynamic textures. Fajita et al. in their article [16] extended the work in [15], and introduced a state variable of an impulse response. Smith et al. in [17] introduced a framework, and represented a video with spatial and temporal wavelets. Otsuka et al. in [18] proposed a method for dynamic textures, based on trajectories of moving contours. Guoying et al. in [22] extended the popular and successful texture based method of local binary patterns (LBP) [20-21], and introduced volume local binary patterns (VLBP) method.

Section 2 talks about the related work from the past literature. Section 3 introduces our proposed framework. Section 4 deals with experimental results. Section 5 concludes the paper.

II. RELATED WORK

In this section, we discuss about some of the existing texture based feature extraction methods. They are local binary patterns, and volume local binary patterns methods.

A. Local Binary Patterns (LBP)

A popular and successful texture based method proposed by Ojala et al. in [20], which has been applied in many fields of research like face recognition, finger print recognition, and image retrieval, etc. The LBP pattern for a 3x3 patch of an image is computed using the following equation (1).

$$LBP(I(CP)) = \sum_{p=1}^{8} 2^{(p-1)} \times f1\left(I(CP) - I(NP_p)\right) \qquad (1)$$

$$Where\ f1(x) = \begin{cases} 1 & x \geq 0 \\ 0 & otherwise \end{cases}$$

In the above equation, CP represents the center pixel. NPp denotes a neighbourhood pixel p in a 3x3 patch of the image, I.

An example LBP value calculation is as shown in figure, Fig.1.

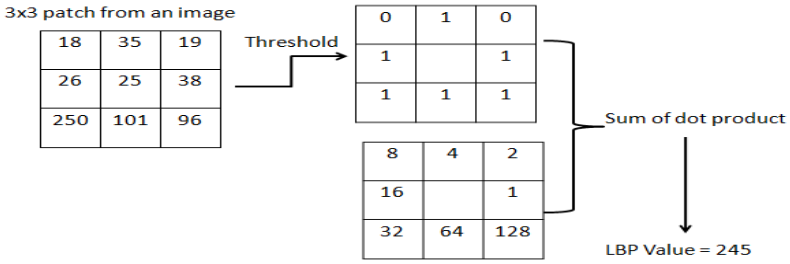


Fig.1. Example LBP calculation

B. Volume Local Binary Patterns (VLBP)

Guoying zhao et al. have proposed the Volume local binary pattern's framework in [22], which is meant for extracting temporal texture features of video. VLBP extends local binary pattern's method. Volume local binary patterns are extracted from the groups of three consecutive frames in a video. From a group of three sequential frames, extract a 3x3x3 cubic patch of pixels, then, threshold the pixel values of the cube using middle frame's center pixel value. The resultant values are multiplied with corresponding binary weights, and summed, which yields the volume local binary patterns as shown in figure, Fig.2.

As shown in Fig.2, the VLBP method depends on four neighbourhoods of the sequence of frames.

III. PROPOSED FRAMEWORK

In this section, we propose our framework. Instead of representing a video using four neighbourhoods based temporal texture features as in VLBP, in our framework, we represent the video using directional temporal local binary patterns (DTLBP).

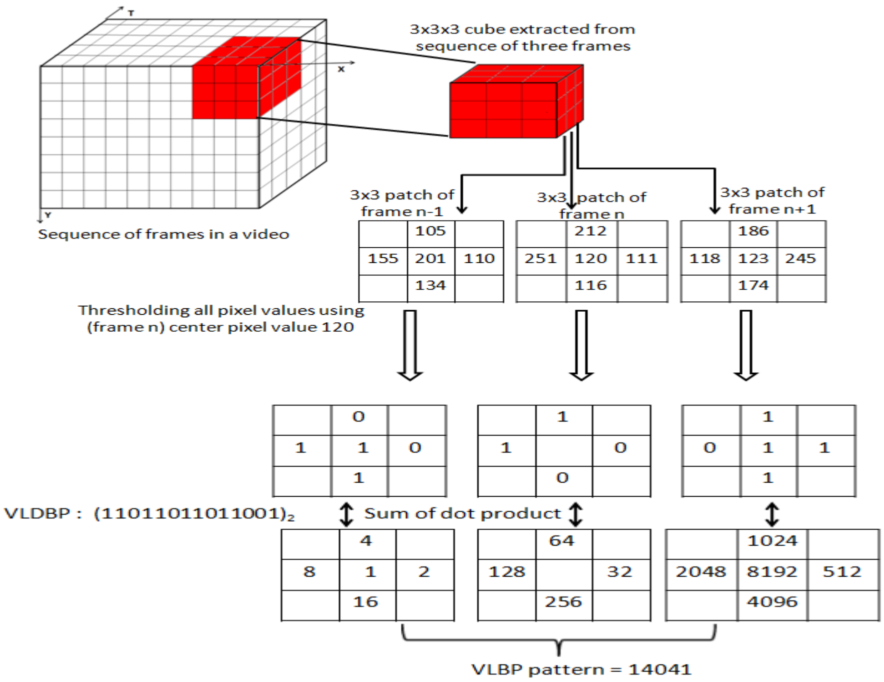


Fig.2. Example VLBP calculation

The figure, Fig.3, illustrates our framework of representing a video with directional temporal local binary patterns. The first step in our framework is, extraction of ten key frames using k-means clustering algorithm. The next step is for each key frame fn forming a group of three frames f_{n-c}, f_n, and f_{n+c}, with a constant time interval of c.

In our case, the c value is fixed to one i.e., if the current frame f_n is 2^{nd} frame then, f_{n-c} is the 1^{st} frame, and f_{n+c} is the 3^{rd} frame, which forms the group of three frames. We follow the same procedure to find the group of frames for each current frame as shown in table 1. Then, extract directional temporal features, for each group of frames using DTLBP method.

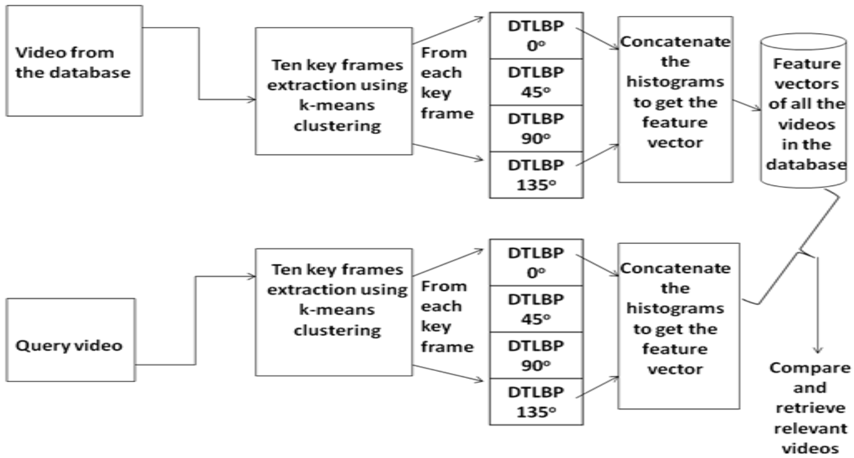


Fig.3. Proposed Video Retrieval Framework

The key frame extraction method consists of following steps:

Step1: From each frame of the input video extract colour histograms.
Step2: The extracted colour histograms in step 1 are grouped in to ten clusters using k-means clustering method.
Step3: Find a frame from each cluster whose colour histogram is nearer to the corresponding cluster center.
Step4: The frame which is selected in step 3 is declared as key frame.

Table 1key Frames And Their Corresponding Group Of Frames

key frames f_n	The corresponding group of three frames		
	f_{n-c}	f_n	f_{n+c}
k_1	k_1-1	k_1	k_1+1
k_2	k_2-1	k_2	k_2+1
k_3	k_3-1	k_3	k_3+1
k_4	k_4-1	k_4	k_4+1
k_5	k_5-1	k_5	k_5+1
k_6	k_6-1	k_6	k_6+1
k_7	k_7-1	k_7	k_7+1
k_8	k_8-1	k_8	k_8+1
k_9	k_9-1	k_9	k_9+1
k_{10}	k_{10}-1	k_{10}	k_{10}+1

C. Directional Temporal Local Binary Patterns (DTLBP)

Our temporal texture feature extraction method is a novel method known as directional temporal local binary patterns (DTLBP), in which, we will extract the temporal features in 0^o, 45^o, 90^o, and 135^o directions. The initial step in DTLBP is, for the current frame f_n, choose two other frames f_{n-c}, and f_{n+c} in the temporal domain with a fixed time interval of c as shown in table 1.

The next step is, from the group of three frames, select a 3x3x3 cube of pixels. From the cube of pixels, extract 3x3 patch of pixels in 0^o direction (highlighted pixels as shown in figure 6(a)) and find the DTLBP value using equation (1). We have followed the same procedure, to extract DTLBP in 45^o, 90^o, and 135^o directions. At last, concatenate all the extracted DTLBPs of four directions. These steps will be applied on all the cube of pixels of the group of three frames, which results in a histogram of DTLBP feature vector. The equation, which is useful for calculating DTLBP is as shown below:

$$DTLBP = cat\left(LBP\left(d_\theta(f_{n-c}, f_n, f_{n+c})\right)\right) \qquad (3)$$

Where the function d_θ (f_{n-c}, f_n, f_{n+c}) selects directional pixels based on θ (θ can have any one of the values 0°, 45°, 90°, and 135°) from the group of three frames (f_{n-c}, f_n, f_{n+c}). And cat represents to concatenate.

An example 3x3x3 cube extracted from the group of three frames is as shown below, in figure, Fig.4.

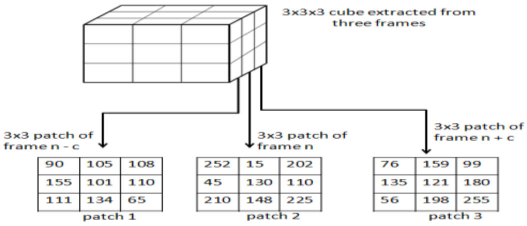


Fig.4. An example 3x3x3 temporal cube of pixels

An example calculation of DTLBP in 0°, 45°, 90°, and 135° directions is as shown in the below figures Fig.5(a), Fig.5(b), Fig.5(c), and Fig.5(d) respectively.

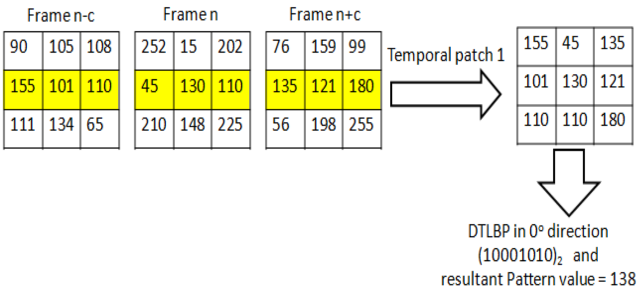


Fig.5 (a) An example calculation of DTLBP in 0° direction

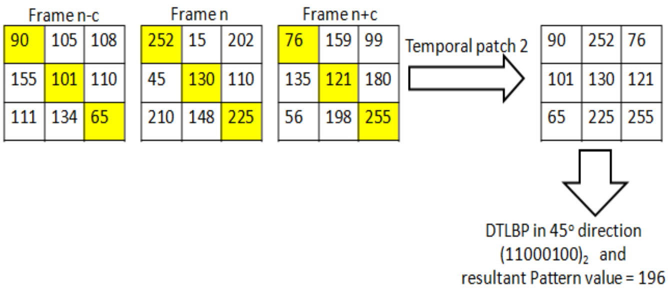


Fig.5 (b) An example calculation of DTLBP in 45° direction

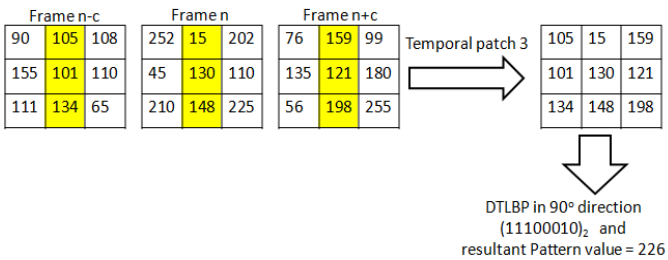


Fig.5 (c) An example calculation of DTLBP in 90° direction

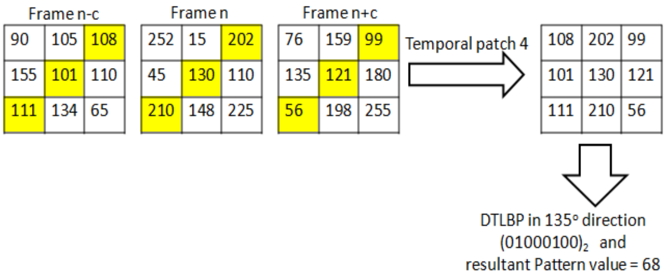


Fig.5 (d) An example calculation of DTLBP in 135° direction

D. Proposed Algorithms

1) **Algorithm 1 - Framework for video indexing and retrieval:**

Step 1. Extract ten key frames from a video using the key frame extraction based on k-means clustering method.
Step 2. For each frame extract directional temporal texture features by using algorithm 2.
Step 3. Concatenate the features extracted in step 2.
Step 4. Apply step 2 for all the selected frames in step 1.
Step 5. Concatenate all the features extracted from above steps to formulate the histogram of feature vector.
Step 6. For a query video, compare the feature vector of the query video with the pre computed feature vectors of the videos in the video database, by using euclidean distance measurement method.
Step 7. Retrieve the relevant videos.

2) **Algorithm 2 – DTLBP extraction algorithm:**

Step 1. For a key frame f_n, from the selected key frames, choose two other frame's f_{n-c} and f_{n+c} in such a way that, there is a fixed time interval c between f_{n-c}, f_n and f_n, f_{n+c}, which forms the group of three frames (In our case c value is one). The groups of frames are as shown in table 1.
Step 2. Find the temporal features for each group of there frames using directional temporal local binary pattern's method, as discussed in section C.
Step 3. Concatenate the extracted temporal texture features of all the four directions 0^o, 45^o, 90^o, and 135^o to formulate the histogram of feature vector for the group of frames.

IV. EXPERIMENTAL RESULTS

We have performed our experiments, on a video data set of three hundred and thirty five videos among which seventy two videos are of boat type; eighty videos are of car type; one hundred and forty eight videos are of airplane type, and thirty five videos are of war tank type. We have compared our proposed frameworks results with already existing method volume local binary patterns (VLBP) method. The retrieval performance is evaluated using precision and recall. Precision and Recall are calculated using following equations (4), and (5).

$$Precision = \frac{true\ positive}{true\ postive + false\ positive} \qquad (4)$$

$$Recall = \frac{true\ positive}{true\ positive + false\ negative} \qquad (5)$$

The results, we have obtained after applying the proposed method DTLBP, and we have compared the results with the already existing method VLBP as shown in table 2, and table 3.

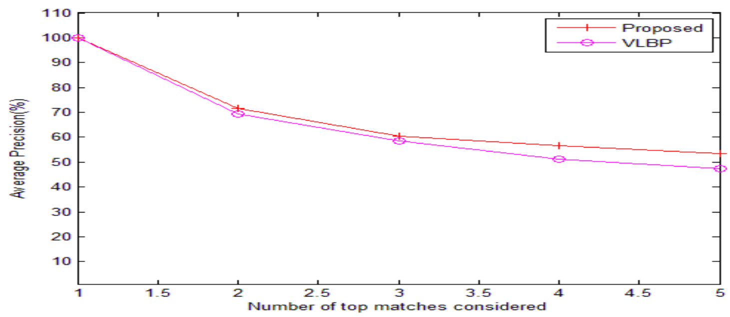
Table 2 Precision (N=5) (%)

Category	**VLBP**	**Proposed DTLBP**
BOATS	46.94	61.39
CARS	47.75	51
AIRPLANES	58.78	67.16
WAR TANKS	36	34.29
AVERAGE VALUE	**47.3675**	**53.46**

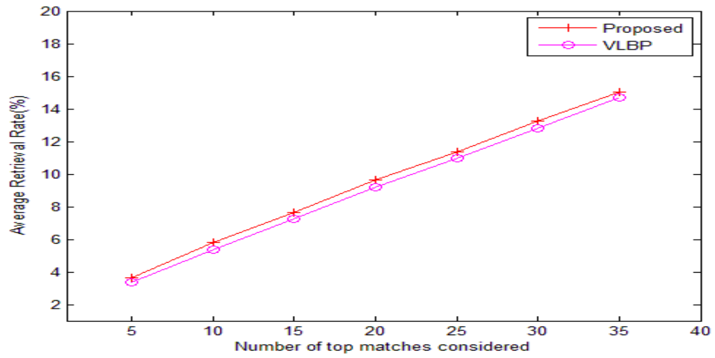
Table 3 Recall (N=35) (%)

Category	**VLBP**	**Proposed DTLBP**
BOATS	16.98	17.57
CARS	15.36	16.77
AIRPLANES	10.57	13.45
WAR TANKS	15.97	12.24
AVERAGE VALUE	**14.72**	**15.0075**

The graphical representation of average performance of our proposed method DTLBP and the existing VLBP method for the different number of top matches retrieved, is as shown in below figures, Fig.7(a), and Fig.7(b).



(a)



(b)

Fig.7. Comparison of proposed method DTLBP with existing method VLBP in terms of (a) average precision (b) average retrieval rate.

By looking at the results, we can say that our proposed algorithm has shown good results with best performance.

V. CONCLUSIONS

In this paper, we have proposed a novel algorithm based on directional temporal local binary patters, which is useful for video indexing and retrieval. We have done our experiments on the data set of three hundred and thirty five videos, and compared the performance of our method with the existing method volume local binary patterns (VLBP). We have found that our algorithm has shown reasonably good results.

REFERENCES

[1]. Szummer.M, Picard.R.W, "Temporal texture modeling". In Proc. IEEE International Conference on Image Processing. Volume 3, (1996) 823-826.

[2]. Doretto.G, Chiuso.A, Soatto.S, Wu Y.N, "Dynamic textures". International Journal of Computer Vision 51(2) (2003) 91-109.

[3]. Peteri. R., Chetverikov.D, "Dynamic texture recognition using normal flow and texture regularity". In Proc. Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA 2005), Estoril, Portugal (2005) 223-230.

[4]. Polana.R, Nelson.R, "Temporal texture and activity recognition". In Motion-based Recognition. Kluwer Academic (1997) 87-115.

[5]. Chetverikov.D, Péteri.R, "A brief survey of dynamic texture description and recognition". In Proc. of 4th Int. Conf. on Computer Recognition Systems. Poland (2005) 17-26.

[6]. Nelson.R.C, Polana.R, "Qualitative recognition of motion using temporal texture". CVGIP: Image Understanding 56 (1992) 78-89.

[7]. Bouthemy.P, Fablet.R, "Motion characterization from temporal co-occurrences of local motion-based measures for video indexing". In Proc. Int. Conf. Pattern Recognition. Volume 1, Brisbane, Australia (1998) 905-908.

[8]. Fablet.R, Bouthemy.P, "Motion recognition using spatio-temporal random walks in sequence of 2D motion-related measurements". In IEEE Int. Conf. on Image Processing, (ICIP 2001). Thessalonique, Greece (2001) 652-655.

[9]. Fablet.R, Bouthemy.P, "Motion recognition using nonparametric image motion models estimated from temporal and multiscale co-occurrence statistics". IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (2003) 1619-1624.

[10]. Lu.Z, Xie.W, Pei.J, Huang.J, "Dynamic texture recognition by spatiotemporal multiresolution histogram". In Proc. IEEE Workshop on Motion and Video Computing (WACV/MOTION'05). Volume 2 (2005) 241-246.

[11]. Peh.C.H, Cheong.L.F, "Exploring video content in extended spatiotemporal textures". In Proc.1st European Workshop on Content-Based Multimedia Indexing. Toulouse, France (1999) 147-153.

[12]. Peh.C.H, Cheong.L.F, "Synergizing spatial and temporal texture". IEEE Transactions on Image Processing 11 (2002) 1179-1191.

[13]. Péteri.R, Chetverikov.D, "Qualitative characterization of dynamic textures for video retrieval". In Proc. International Conference on Computer Vision and Graphics (ICCVG 2004). Warsaw, Poland (2004).

[14]. Fazekas.S, Chetverikov.D, "Normal versus complete flow in dynamic texture recognition: a comparative study". Texture 2005: 4th International Workshop on Texture Analysis and Synthesis, Beijing (2005). http://visual.ipan.sztaki.hu/publ/texture2005.pdf.

[15]. Saisan.P, Doretto.G, Wu.Y.N, Soatto.S, "Dynamic texture recognition". In Proceedings of the Conference on Computer Vision and Pattern Recognition. Volume 2, Kauai, Hawaii (2001) 58-63.

[16]. Fujita.K, Nayar.S.K, "Recognition of dynamic textures using impulse responses of state variables". In Proc. Third International Workshop on Texture Analysis and Synthesis (Texture 2003). Nice, France (2003) 31-36.

[17]. Smith.J.R, Lin.C.Y, Naphade.M, "Video texture indexing using spatiotemporal wavelets". In IEEE Int. Conf. on Image Processing (ICIP 2002). Volume 2 (2002) 437-440.

[18]. Otsuka.K, Horikoshi.T, Suzuki.S, Fujii.M, "Feature extraction of temporal texture based on spatiotemporal motion trajectory". In ICPR. Volume 2 (1998) 1047-1051.

[19]. Zhong.J, Scarlaro.S, "Temporal texture recognition model using 3D features". Technical report, MIT Media Lab Perceptual Computing (2002).

[20]. Ojala. T, Pietikainen.M, Harwood.D, "A comparative study of texture measures with classification based on feature distributions". Pattern Recognition 29 (1996) 51-59.

[21]. Ojala.T, Pietikainen.M, Maenpaa.T, "Multiresolution gray scale and rotation invariant texture analysis with local binary patterns". IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7) (2002) 971-987.

[22]. Guoying Zhao, Matti Pietikainen, "Dynamic Texture Recognition Using Volume Local Binary Patterns".

M.Ravinder has completed his B.Tech, and M.Tech in CSE, in 2003 and 2009 respectively. He has more than eight years of teaching experience and currently pursuing PhD in CSE from JNTUK, Kakinada, AP, India.

Dr.T.Venugopal has completed his M.Tech, and PhD in CSE, in 2003 and 2009 respectively. He has more than seventeen years of teaching experience and currently he is working as Associate Professor in CSE at JNTUHCES, Sultanpur, Medak, Telangana, India.