Compressed domain zoom motion detection and classification based on application of local ternary patterns on block motion vectors

Pavan Sandula, Manish Okade
Intelligent Systems Lab, Dept. of ECE, National Institute of Technology Rourkela
Rourkela, Odisha, India
516EC6004@nitrkl.ac.in, okadem@nitrkl.ac.in
https://sites.google.com/site/manishokade/publications

MOTIVATION & PROBLEM FORMULATION

Motivated by local patterns which essentially capture the image structure using texture analysis and have been previously utilized in many vision based applications like face recognition, content base image retrieval to name a few

We investigated whether this concept can be applied to the compressed domain block motion vectors with the objective of recognizing and classifying the zooming frames occurring in video sequences.

Problem Formulation:

- Our idea is to analyse this orientation field on similar terms as done with traditional pixel intensities in zoom motion detection upon a block motion vector field.
- To this effect we would need two vital aspects. Firstly, the choice of size of neighborhood which in our case is set to 3×3 and secondly, the knowledge of orientation patterns in case of zooming and non-zooming block motion fields.
- The choice of encoding texture descriptor must account for this constraint which makes LTP a natural fit to identify zooming frames and its types.

PROPOSED METHOD

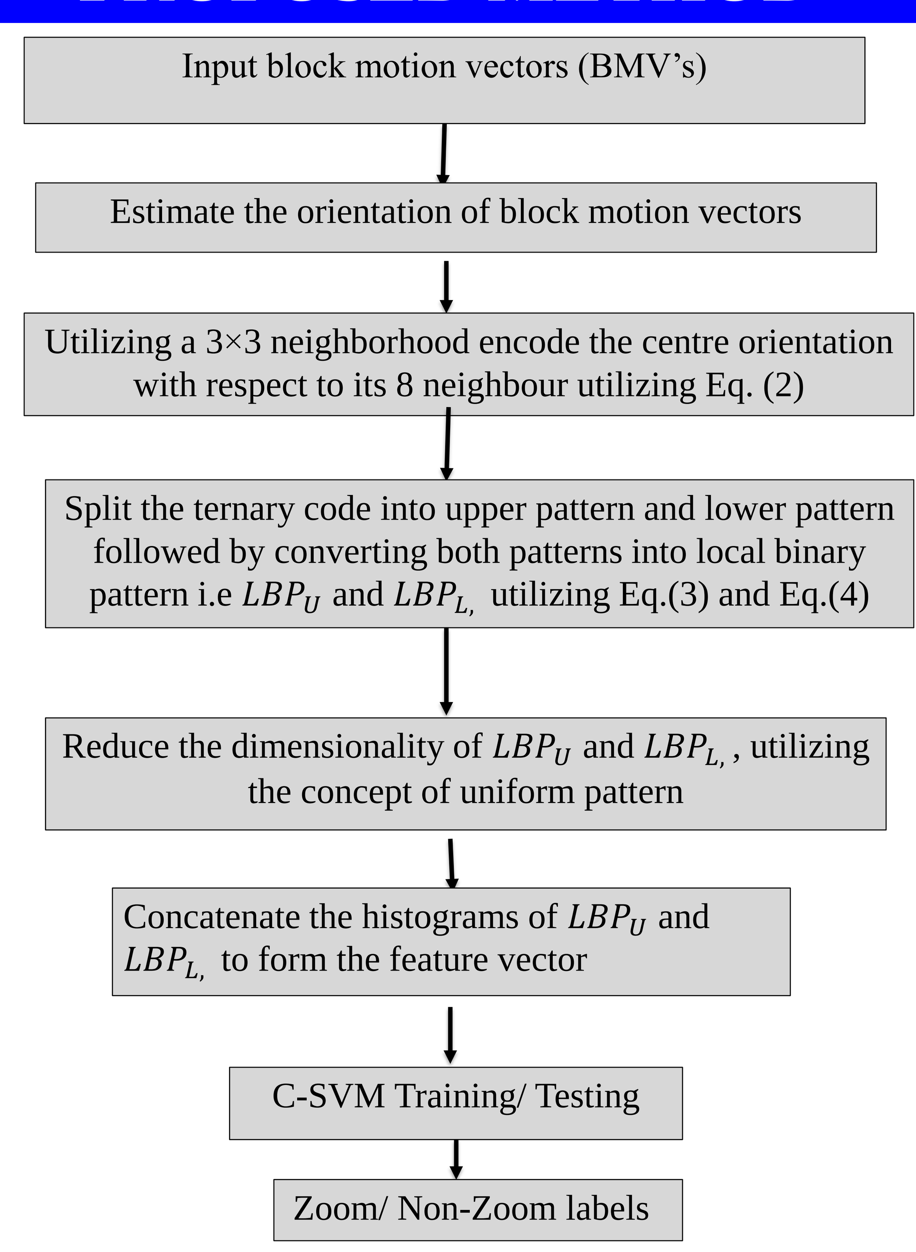

Fig1. Overview of proposed zoom motion detection system

$$MV_{ori} = \arctan\left(\frac{MV^Y}{MV^X}\right) \qquad (1)$$

$$S(MV_{ori}^i, MV_{ori}^c, t) = \begin{cases} 1 & MV_{ori}^i \geq MV_{ori}^c + t \\ 0 & |MV_{ori}^i - MV_{ori}^c| < t \\ -1 & MV_{ori}^i \leq MV_{ori}^c - t \end{cases} \qquad (2)$$

$$LBP_U(x_c, y_c) = \sum_{n=0}^{7} 2^n.(upper\ pattern) \qquad (3)$$

$$LBP_L(x_c, y_c) = \sum_{n=0}^{7} 2^n.(Lower\ pattern) \qquad (4)$$

Dimensionality reduction using uniform patterns

- Since 8 neighborhood is considered, 2^8 (256) combinations each is possible for the upper and lower pattern. This increase the computational cost of feature vector formed (256×2) using the two patterns. In order to reduce this we use the concept of uniform patterns where the pattern is checked for at most one '0-1' and one '1-0' transition where viewed as circular bit string.
- Such patterns characterize the important structural information and must be retained and the rest referred as non-uniform can be discarded since they contain very less information. In case of 'P' neighborhood, [0,…P(P-1)+2] patterns are uniform.
- The histograms corresponding to LBP_U and LBP_L are formed separately followed by discarding the non-uniform patterns thereby forming 59 bin feature vector (0 to 58) out of 256 patterns are uniform.
- The histograms corresponding to LBP_U and LBP_L are formed separately followed by discarding the non-uniform patterns thereby forming 59 bin feature vector (0 to 58) for upper pattern and lower pattern respectively as given below.

$$H_{LBP_\alpha}(l) = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} f_1\,(LBP_\alpha, l)$$

where, $l \in$ [0 to 58], LBP_α is the local binary pattern with $\alpha\epsilon$ (U,L) and $\alpha\epsilon(U, L)$ and $M \times N$ is the size of the motion vector orientation field and

$$f_1(x,y) = \begin{cases} 1\ if\ x = y \\ 0, elsewhere \end{cases}$$

The feature vector is formed by concatenating the histograms corresponding to the upper and lower LBP and is of size 118 (59 × 2) i.e,

$$F.V. = [H_{LBP_U}, H_{LBP_L}]$$

Zoom-in and zoom-out classification:

The objective in this section is to separate the zooming frames which have identified in the earlier section into zooming-in camera type and zoom-out camera type. Zoom-in camera is one where the environment under capture is brought nearer to the camera which causes the motion vectors to diverge from the centre of frame while zoom-out is one where the environment under capture is brought nearer to the camera which causes the motion vector to diverge from the centre of frame

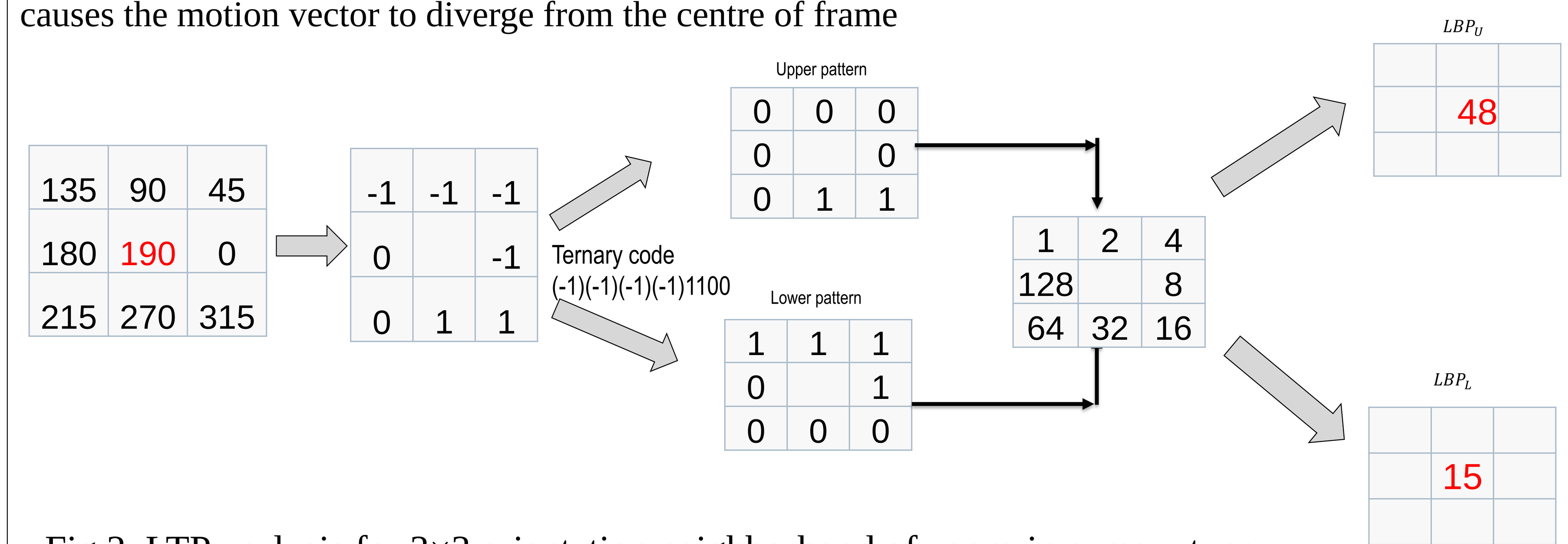

Fig 2. LTP analysis for 3×3 orientation neighborhood of zoom-in camera type

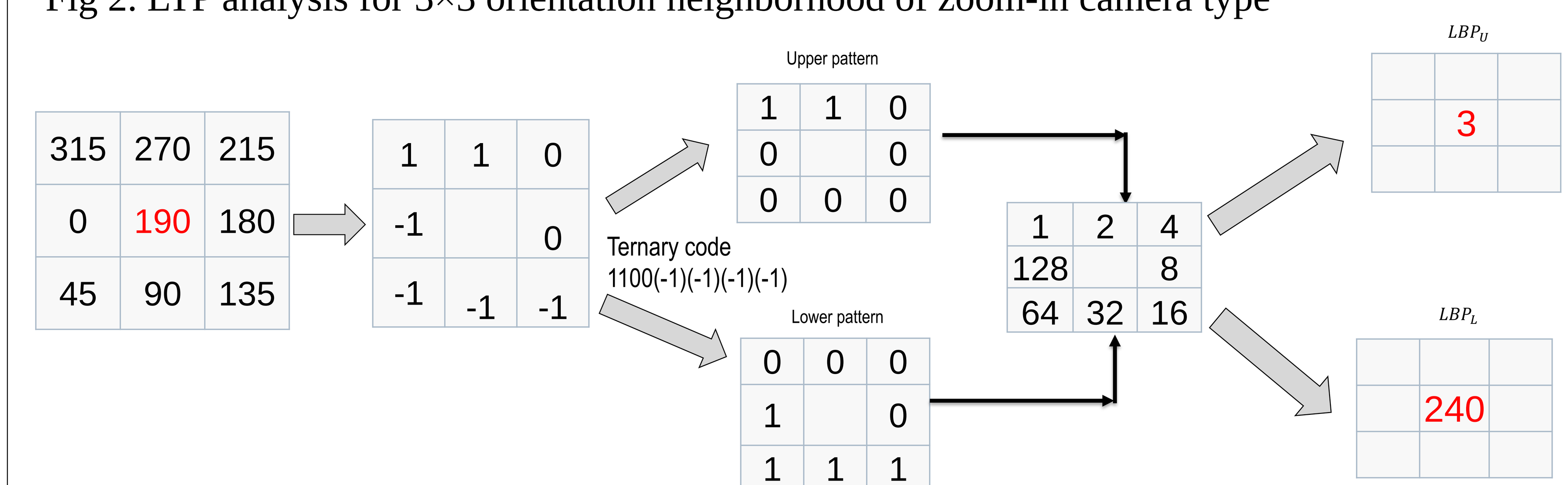

Fig3. LTP analysis for 3×3 orientation neighborhood of zoom-out camera type

$$F.V. = \frac{H_{LBP_U}}{H_{LBP_L}}$$

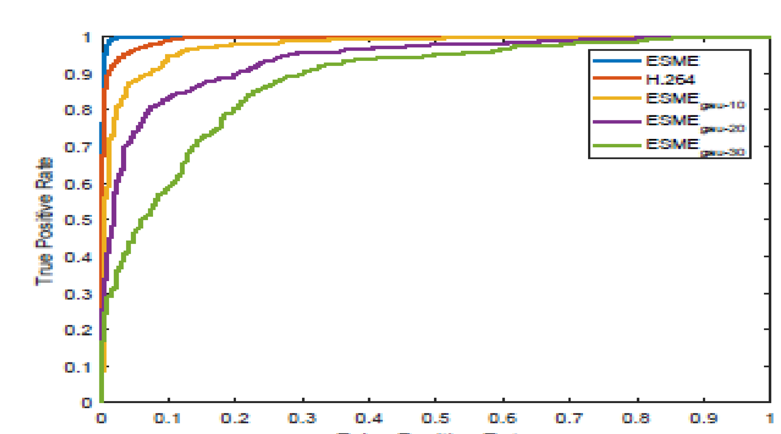

Fig 4. ROC curves demonstrating the zoom detection performance on various block motion vector types

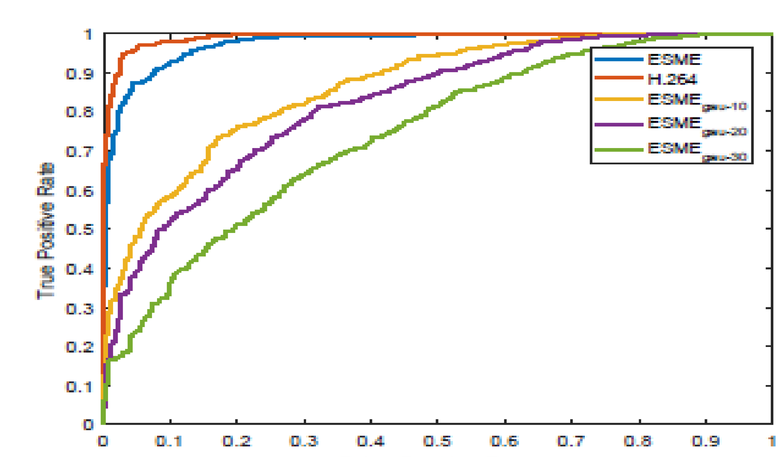

Fig 5. ROC curves demonstrating the zoom motion classification performance on various block motion vector types

Block Motion Vector type	Proposed method
ESME	0.9995
ESME corrupted with Gaussian noise($\sigma^2 = 10$)	0.9785
ESME corrupted with Gaussian noise ($\sigma^2 = 20$)	0.9403
ESME corrupted with Gaussian noise($\sigma^2 = 30$)	0.8798
H.264	0.9959

Area under Curve (AUC) for zoom motion detection demonstrating the performance on various block motion vectors

Block Motion Vector Type	Accuracy (%)			
	Parametric method	Duan et al. method	Okade et al method	Proposed method
ESME	53.87	91.08	92.25	98.81
ESME corrupted with Gaussian noise ($\sigma^2 = 10$)	50.25	57.41	51.01	84.98
ESME corrupted with Gaussian noise ($\sigma^2 = 20$)	48.72	51.25	50.16	73.34
ESME corrupted with Gaussian noise ($\sigma^2 = 30$)	47.45	50.41	49.83	66.58
H.264	56.84	81.53	94.81	95.39

Accuracy (%) for zoom motion detection at false positive rate set to 1%

Block Motion Vector type	Proposed method
ESME	0.9755
ESME corrupted with Gaussian noise($\sigma^2 = 10$)	0.8631
ESME corrupted with Gaussian noise ($\sigma^2 = 20$)	0.8238
ESME corrupted with Gaussian noise($\sigma^2 = 30$)	0.7426
H.264	0.9916

Area under Curve (AUC) for zoom motion classification demonstrating the performance on various block motion vectors

Block Motion Vector Type	Accuracy (%)		
	Parametric method	Duan et al. method	Proposed method
ESME	53.87	76.03	84.08
ESME corrupted with Gaussian noise ($\sigma^2 = 10$)	50.25	57.41	64.91
ESME corrupted with Gaussian noise ($\sigma^2 = 20$)	48.72	51.25	59.25
ESME corrupted with Gaussian noise ($\sigma^2 = 30$)	47.45	50.41	57.75
H.264	56.84	81.53	90.31

Accuracy (%) for zoom motion classification at false positive rate set to 1%

CONCLUSION

- This paper investigated the application of local ternary pattern which was earlier used for image texture analysis to the zoom motion detection and classification problem in video sequences.
- A 3 × 3 neighborhood was utilized for this purpose by encoding the eight orientation values with respect to the centre orientation value.
- The ternary code so obtained was split into its upper and lower patterns for which binary patterns were estimated.
- Uniform pattern based dimensionality reduction was carried out followed by the formation of the feature vector which was utilized to train the C-SVM classifier for zoom motion detection and its ratio for zoom motion classification.
- Experimental validation carried out utilizing ESME and H.264 obtain block motion vectors extracted from standard video sequences

REFERENCES

1. L-Y Duan, J.S. Jin, Q. Tian and C-S. Xu, " Nonparametric motion characterization for robust classification based of camera motion parameters," IEEE transaction on multimedia, vol.8, pp. 323-340, Apr. 2006.
2. M Okade, G.patel and P.K.Biswas, "Robust learning based camera motion characterization scheme with application to video stabilization," *IEEE Transactions on Circuits and Systems for video technology*, Vol.26,pp.453-466 , Mar.2016.
3. M. Guironnet, D. Pellerin and M. Rombaut, "camera motion classification based on transferable belief model," in 14th European Signal Processing conference,"(Florence, Italy), pp. 1-5, IEEE, Sep 2006.
4. X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," IEEE Transactions on Image Processing, vol. 19,pp. 1635–1650, June 2010.