International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 17 (2017) pp. 6897-6900

Comparative performance analysis of LBP and LTP based Facial Expression Recognition

Smt. Shaik Taj Mahaboob
Assistant Professor, Department of Electronics and Communication Engineering, Jawaharlal Nehru Technological University Anantapur college of Engineering, Muddanuru road, Pulivendula, Andhra Pradesh- 516390, India.

Orcid Id: 0000-0002-7447-4669

Dr. S. Narayana Reddy
Professor, Department of Electronics and Communication Engineering, Sri Venkateswara University College of Engineering, Sri Venkateswara University, Tirupathi, Andhra Pradesh, India.

Abstract

Local Binary Patterns (LBP) are extensively used for facial expression classification. Several variants of LBP have been proposed, e.g. Local Ternary Patterns (LTP) to make LBP resilient to noise, Scale Invariant LTP (SILTP) to make LBP resilient to illumination changes. But neither LTP nor SILTP are resilient to both noise and illumination changes. This chapter proposes a generalised variant of LBP called the Generalised Local Ternary Patterns (gLTP) which captures edge and blob-like features and makes the LBP resilient to both noise and illumination changes. Experiments on two datasets (Facial Expression Recognition) show that neither LTP nor SILTP gives better performance on either dataset. On the other hand, the proposed gLTP descriptor gives competitive performance compared to the best performing descriptors in the datasets, confirming that gLTP is resilient to noise and illumination.

Keywords: LBP, LTP, Scale Invariant LTP, Generalized LTP, Expression Classification, Facial Expression Recognition, Descriptors.

INTRODUCTION

Local Binary Patterns, proposed by Ojala et al. [1], have proved a very powerful texture descriptor which has been widely applied for e.g. texture classification [2], face recognition [3], medical image classification [4, 5]. LBP describes the local texture around each pixel by comparing and thresholding pixel differences in a local image neighborhood. A global image representation of an image is normally obtained by computing statistical representations (e.g. histogram) of the LBP-based pixel representations.

Several variations of LBP have been proposed, e.g. Local Ternary Patterns (LTP) [15] makes LBP resilient to noise, Scale Invariant LTP (SILTP) [9] makes LBP resilient to illumination changes, uniform LBP [7] captures informative patterns such as edges, bright and dark spots, and reduces the dimensionality of the histogram image representation, Block LBP[6] captures information from larger local neighborhood (e.g. larger than 3x3 which is often used by LBP). In the following sections first LBP and its major variants will be concisely reviewed, then the proposed gLTP descriptor described in detail, and finally several experiments comparing gLTP with baseline representations reported.

LBP AND ITS VARIANTS

The standard LBP descriptor

Consider N sampling points distributed uniformly on a circle of radius R around a 2D point $\mathbf{p}_c$ in a gray image I (Figure 1). LBP can be defined as:

$$LBP_{N,R}(p_c) = \sum_{n=1}^{N} qn \times 2^{n-1} \text{ where } qn = \begin{cases} 1 & In \geq \\ 0 & In \leq \end{cases} \quad \text{... Eq 1}$$

Ic and In represent the intensity values at the center point ($\mathbf{p}c$) and at the n-th sampled image point, respectively. In is bilinearly interpolated when the sampling point does not coincide with a pixel. Since this operator gives $2N$ different labels, an image can be represented as a histogram with $2N$ bins.

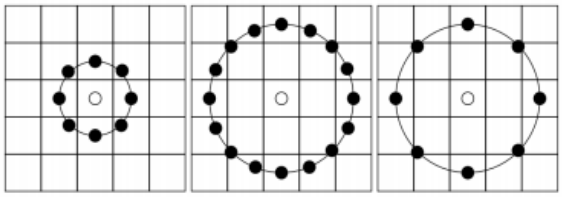


Figure 1: The circular (8, 1), (16, 2) and (8, 2) neighborhoods.

The pixel values are bi-linearly interpolated whenever the sampling point is not in the center of a pixel. The generation of the standard LBP codes from a 8 neighborhood is illustrated in Figure 2. and Figure 3 shows an example image, its corresponding LBP representation.

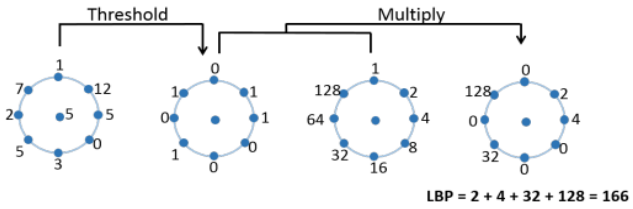
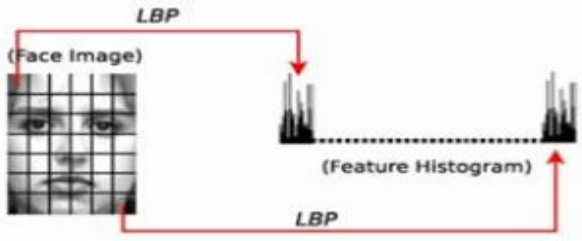
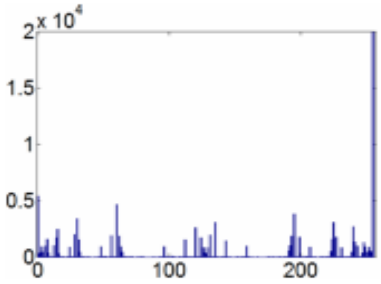


Figure 2: Example illustrating the derivation of the standard LBP codes. The pixels in this block are threshold by its center pixel value, multiplied by powers of two and then summed to obtain a value for the center pixel.



(i) original image



(ii) LBP histogram

Figure 3. A face image is divided into small regions from which LBP histograms are extracted and then concatenated into a single, spatially enhanced feature histogram.

Uniform and rotation invariant LBP

As some of the binary patterns occur more commonly in texture images than others, an extension of LBP, called the *uniform LBP* [8], has been proposed. The uniform patterns describe frequently occurring basic features such as bright spots, dark spots and edges. A LBP is called uniform if the binary pattern (Equation1) contains at most two bitwise transitions from 0 to 1 or vice versa when the bit pattern is traversed circularly.

For example, the patterns 00000000 (0 transitions), 01110000 (2 transitions) and 11001111 (2 transitions) are uniform whereas the patterns 11001001 (4 transitions) and 01010010 (6 transitions) are not. Figure 4 shows some examples of uniform and non-uniform patterns. The resultant LBP representation has a separate label for each uniform pattern and all the non-uniform patterns are assigned to a single label. For example, when N = 8 the number of patterns produced by the uniform LBP is 59 compared to the total number of 256 patterns produced by the standard LBP ([10]).

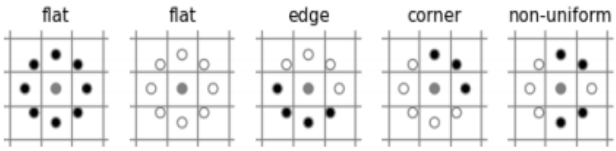


Figure 4: Some example uniform (first four images) and non-uniform (last image) patterns.

Uniform patterns have several advantages: (1) they capture more commonly occurring local structures, (2) considering the uniform patterns makes the number of possible LBP labels significantly lower, hence require fewer samples to estimate their distribution reliably and (3) the dimensionality of the final image representation is reduced, hence reducing the classification complexity. It has been observed that considering only the uniform patterns instead of all the possible patterns produces better recognition results for many applications [11].

Rotation-invariant LBP makes the LBP descriptors invariant to **local image region** rotations by rotating the LBP binary codes in a circular bit-wise manner so that the resultant LBP label will have the minimum value [12]

Local Ternary Patterns

To make LBP robust to noise, a three-level thresholding has been applied in *Local Ternary Patterns (LTP)* [13] by the introduction of a user-specified threshold τ Equation (2). The LTP histogram representation of an image is obtained by splitting each LTP into two LBP and then concatenating the two LBP-based representations.

$$q_n(\tau) = \begin{cases} 1 & I_n - I_c \geq \tau \\ 0 & |I_n - I_c| < \tau \\ -1 & I_n - I_c \leq -\tau. \end{cases}$$

…. Eq 2

An ideal LBP should be robust to illumination changes and (at least) Gaussian noise. While the introduction of τ makes LTP robust to noise, LTP may sensitive to changes in illumination Figure.5

Scale Invariant Local Ternary Patterns

To counteract illumination variations, a variant of LTP called the *Scale*1 *Invariant Local Ternary Pattern* (SILTP) has been proposed in [14], i.e

$$SILTP_{N,R}(\mathbf{p}_c, a) = \oplus_{n=1}^{N} q_n(a)$$

$$\text{where} \quad q_n(a) = \begin{cases} 01 & I_n > (1+a)I_c \\ 10 & I_n < (1-a)I_c \\ 00 & \text{otherwise,} \end{cases}$$

… **Eq.3**

where a is a scale factor and $\oplus$ denotes concatenation of the 2-bit binary strings qn. Note that SILTP is not designed particularly for image classification, but the '2-bit' codes can be converted to 'ternary' patterns to generate a histogram representation for an image. Since SILTP is designed to cope with the changes in illumination it may sensitive to noise from Figure 5.

Other variants

Inspired by the success of LBP in various computer vision applications, different variants of LBP have been proposed to increase robustness and discriminative power. Since LBP uses zero as the threshold to compare a pixel with its neighborhood, several alternative thresholding techniques have been proposed, e.g. median and mean of the local neighborhood is used as the threshold in [4] and [2] respectively. Usually LBP operates on a small image neighborhood (3 × 3). To capture larger image neighborhoods Gaussian filtering is applied to collect intensity information from an area larger than the original single pixel in [13], the averaged pixel values in small image blocks were used in [5]. I described only the major relevant variants. A complete review on LBP variants can be found in [13].

Generalised Local Ternary Patterns

LTP and SILTP make the LBP representations resilient to noise and illumination changes respectively. But neither LTP nor SILTP are resilient to both noise and illumination changes. Therefore, we propose a generalized variant of LBP called the Generalized Local Ternary Patterns (gLTP) which makes the LBP resilient to both noise and illumination changes

Definition

When a scene is illuminated by a single distant light source, the observed luminance image $I(x, y)$ at point (x, y) can be approximated as the product of the reflectance image $R(x, y)$ and the illuminance image $S(x, y)$ [15], i.e.

$$I(x, y) = S(x, y)R(x, y) + G(x, y) \text{.....} \quad \textbf{Eq.4}$$

Consider two pixels $I(x1, y1)$ and $I(x2, y2)$ in image I, and the difference $D = I(x1, y1) - I(x2, y2)$. Under a different illumination, using Equation-4 D becomes:

$$D = [a1I(x1, y1) + \tau 1] - [a2I(x2, y2) + \tau 2] \propto I(x1, y1) - aI(x2, y2) - \tau.$$

..... **Eq.5**

where $a1$ and $a2$ represent the non-uniform illumination applied to the pixels $I(x1, y1)$ and $I(x2, y2)$, $\tau 1$ and $\tau 2$ are the sensor noise due to the image capturing device at those pixels, and $a = a2/a1$, $\tau = \tau 2 - \tau 1/\ a1$.

From Equation.5 we can observe that LBP is not resilient to noise (Figure 5(ii)), as it assumes $\tau = 0$. LTP is not robust to illumination changes (Figure 5(iii)) as it considers $a = 1$. SILTP is robust to illumination changes but it assumes the noise dependent on pixel values (SILTP can be rewritten as, e.g. $qn(a) = 01$ when $In - aI\ c > \tau$, where $\tau = Ic$). To make the LBP robust to noise *and* illumination changes, my formulation considers $a \in \mathrm{R}$ and $\tau \geq 0$ (Gaussian noise). The proposed formulation *generalized Local Ternary Patterns* (gLTP), becomes:

$$q_n(a, \tau) = \begin{cases} 1 & I_n - aI_c \geq \tau \\ 0 & |I_n - aI_c| < \tau \\ -1 & I_n - aI_c \leq -\tau. \end{cases}$$

…..Eq. 6

The standard LBP ($a = 1$, $\tau = 0$), LTP ($a = 1$, $\tau > 0$) and the SILTP ($a \in \mathrm{R}$, $\tau = Ic$) can be seen as special cases of gLTP. The proposed formulation outputs ternary patterns; I convert each ternary pattern into two binary patterns as in the standard LTP

Table 1: The gLTP with different parameter settings (Equation (6)). LBP is not resilient to noise, as it assumes $\tau = 0$. LTP is not robust to illumination changes as it considers $a = 1$. SILTP is robust to illumination changes but it assumes the noise dependent on the value of the center pixel. gLTP is a generalization of LBP, LTP and SILTP and robust to *both* noise *and* illumination changes.

Parameters	LBP	LTP	SILTP	gLTP
a	1	1	$\in \mathrm{R}$	$\in \mathrm{R}$
τ	0	$\in \mathrm{R}$	Ic	$\in \mathrm{R}$

Figure 5 shows an image patch, its three transformed versions (changed illumination, noise and both) and its LBP, LTP, SILTP and gLTP codes. This figure shows that LTP is resilient to noise but not to illumination changes and SILTP is resilient to illumination changes but not to noise. Changing illumination and noise (last row, Figure 5), gLTP yields the most stable output compared to the other methods.

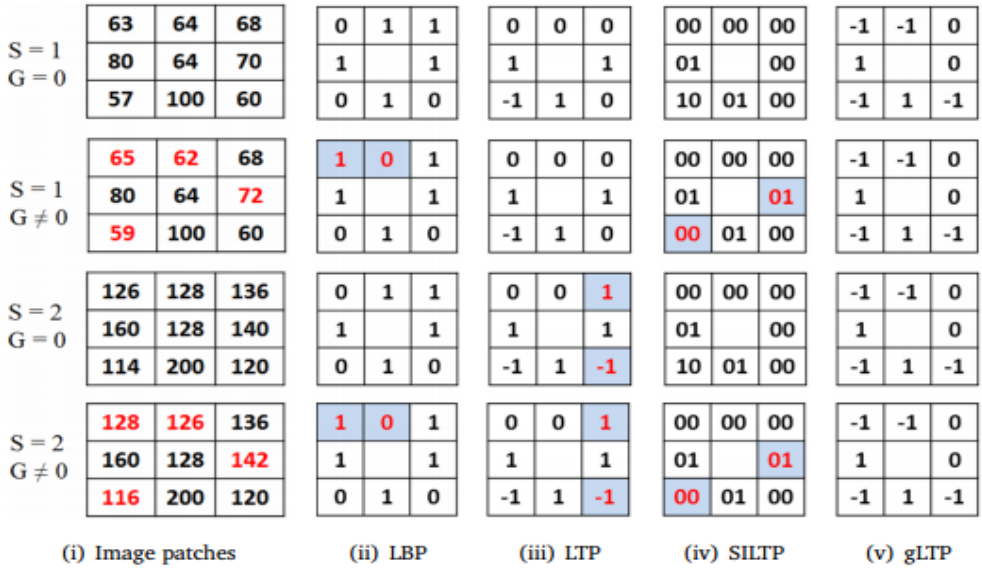


Figure 5: effect of noise and illumination changes

Effect of parameters

Here it is shown qualitatively that the gLTP can capture edge and blob-like features in an image by appropriate parameter setting. The first row of Figure.6 shows a CK-dataset image from a standard in-air procedure and its LBP, LTP and gLTP codes. The remaining rows show the original image under different illumination and noisy conditions, and the codes computed by LBP, LTP, SILTP and gLTP. To mimic the spot illumination which is often used in CK-dataset procedure, first an un-normalized Gaussian filter was created with a window size which is equal to 3*w* and a standard deviation equals to 2*w* / 3, where *w* is the width of the image. Then randomly select a point in the image and placed this Gaussian filter. The pixel values of the image are then multiplied by this filter, and the resultant values are then clipped at 255 to make sure they are in [0,255].

A demonstrative example for the effect of noise and illumination changes on (b) LBP, (c) LTP (τ = 5), (d) SILTP (w = 0.1), and (e) the proposed gLTP (w = 0.9, τ = 5). (a) First row: original image patch with 3 × 3 pixels (i.e. s = 1 and G = 0 in Equation 4); Second row: noise added to the original image patch (s = 1 and G 6= 0) shown in red; Third row: the original image patch under a different illumination (s = 2 and G = 0); Fourth row: the original image patch under a different illumination with noise added (s = 2 and G 6= 0) (noisy pixels are shown in red); LTP is robust to noise but not to illumination changes. LBP and SILTP are robust to illumination but not to noise. LBP, LTP and SILTP are sensitive to both illumination and noise (last row). The proposed gLTP is robust to both noise and illumination. It is clear that LBP, LTP and SILTP (second, third and fourth columns) gives different output codes under different conditions (noise, illumination or both). On the other hand, the codes generated by gLTP capture edge-like features and are less affected by illumination and noise transformations. Figure7 shows an example histological image and the resulting LBP, LTP, SILTP and gLTP codes. Since the original cell image itself is very noisy, LBP and SILTP gives very different outputs under different conditions. gLTP captures blob-like features and is less affected by the illumination and/or noise transformations.

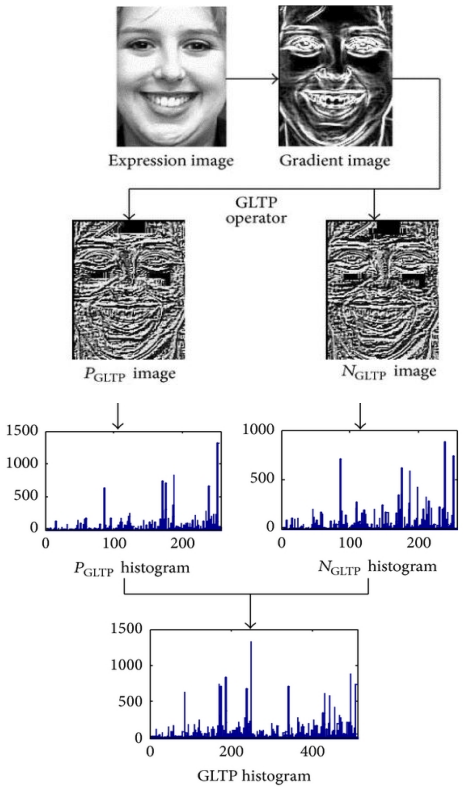


Figure 6. gLTP Histograms

Spatial histograms computed from the whole encoded image do not reflect the location information of the micropatterns, only their occurrence frequencies are represented [1]. However, it is understandable that a histogram representation that combines the location information of the GLTP micro-patterns with their occurrence frequencies is able to describe the local texture more accurately and effectively [2, 3]. Therefore, in order to incorporate some degree of location information with the GLTP histogram, each facial image is divided into a number of regions, and individual GLTP histograms (representing the occurrence information of the micro-patterns from the corresponding local region) computed from each of the regions are concatenated to obtain a spatially combined GLTP histogram. In the facial expression recognition system, this combined GLTP histogram is used as the facial feature vector. The process of generating the combined GLTP histogram is illustrated in Figure 6.

EXPERIMENTS AND RESULTS

Experimental Setup and Dataset Description

To evaluate the effectiveness of the proposed face feature descriptor, experiments were conducted on images collected from a well-known image database, namely, the Cohn-Kanade (CK) facial expression database [1]. In the CK database, a sample set of 100 students, aging from 18 to 30 during image acquisition, were included. A majority of the subjects (65%) were female; 15% of the samples were African-American, and 3% were Asian or of Latin descent. Each of the students displayed facial expressions starting from non-expressiveness to one of the aforementioned six prototypic emotional expressions in the image acquisition process. These image sequences were then digitized into 640 × 480 or 640 × 690 pixel resolutions. In our setup, a set of 1224 facial image sequences were selected from 96 subjects, and each of the images was given a label describing the subject's facial expression. The dataset containing the 6 classes of expressions was then extended by 408 images of neutral facial images to obtain the 7-class expression dataset. Figure 7 shows sample prototypic expression images from the CK database.

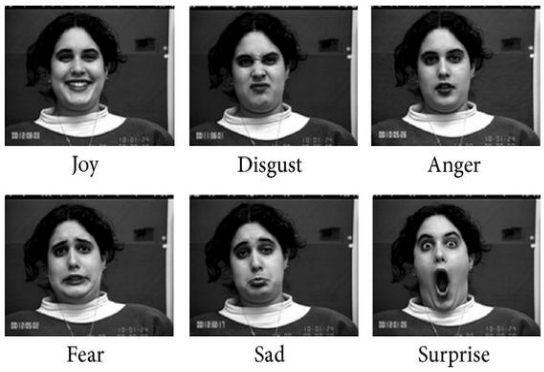


Figure 7. Sample 6-class expression images from the CK database

We cropped the selected images from the original ones based on the ground truth of the positions of two eyes, which were then normalized to 150 x110 pixels. Figure 8 shows a sample cropped facial image from CK database. A tenfold cross-validation was carried out to compute the classification rate of the proposed method. In tenfold cross-validation, ten subsets comprising equal number of instances are formed by partitioning the whole dataset randomly. The classifier is first trained on the nine subsets, and then the remaining set is used for testing. This process is repeated for 10 times, and the average classification rate is computed. The threshold value t was set to 10 empirically.

Figure 8. Cropping of a sample face image from the original one

EXPERIMENTAL RESULTS

The classification rate of the proposed method can be influenced by adjusting the number of regions into which the expression images are to be split [2]. We have considered three cases in our experiments as opted in [2], where images were divided into, and regions. We have compared our proposed method with 3 widely used local texture descriptors, namely, local binary pattern (LBP) [16], local ternary pattern (LTP) [4], and local directional pattern (LDP) [2]. Tables 1 and 2 show the classification rates of these local texture descriptors for the 6-class and the 7-class expression recognition problem, respectively. It can be observed that dividing an image with higher number of regions will produce higher classification rate, since the feature descriptor then contains more location and spatial information of the local patterns. However, the feature vector length will also be higher in such cases, which affects the computational efficiency. Hence, selection of the number of regions is a trade-off between computational efficiency and classification rate.

Table 2: Recognition rate (%) for the CK 6-class expression dataset using different local texture descriptors

Operator	Classification rate (%) for different number of regions		
	3x3	5x5	7x6
LBP	79.3	89.7	90.1
LTP	87.3	92.3	93.6
gLTP	90.5	96.4	97.2

For both the 6-class and the 7-class expression recognition problems, the proposed GLTP feature descriptor achieves the highest recognition rate for images partitioned into different number of regions. For the 6-class dataset, GLTP achieves an excellent recognition rate of 97.2%. On the other hand, for the 7-class dataset, the recognition rate is 91.7%. Here, inclusion of neutral expression images results in a decrease in the accuracy. For both the 6-class and the 7-class recognition problems, the highest classification rate is obtained for images partitioned into 7x6 regions.

Table 3: Confusion matrix of CK 6-class recognition using GLTP for images partitioned into 7x6 regions

	Anger (%)	Disgust (%)	Fear (%)	Joy (%)	Sad (%)	Surprise (%)
Anger	98.4	0	0	0.8	0	0.8
Disgust	0.5	94.4	0	0	5.1	0
Fear	0	0.7	97.1	0	4,8	1.8
Joy	1.1	1.1	0	97.8	0	0
Sad	0	4.5	0	0	95.5	0
Surprise	0	0	0	0	0	100

Table 4. Parameters

LBP variant	a	τ
LTP	-	[5, 10, 20, 30, 40]
SILTP	[0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 0.9]	-
gLTP	[0.95, 0.9, 0.8, 0.7, 0.5, 0.3, 0.1]	[5, 10, 20, 30, 40]

The confusion matrix of recognition using the gLTP descriptor for the 6-class and the 7-class datasets is shown in Tables 3 and 4, respectively, which provides a better picture of the recognition accuracy of individual expression types. It can be observed that, for the 6-class recognition, all the expressions can be recognized with high accuracy. For the 7-class dataset, while anger, disgust, fear, joy, and surprise can be recognized with high accuracy, the recognition rates of sadness and neutral expressions are lower than the average. Evidently, inclusion of neutral expression images results in a decrease in the accuracy, since many sad expression images are confused with the neutral expression images and vice versa.

Parameter selection

At each iteration of the experiment I apply a 3-fold cross validation on the training set to select the parameters for LTP, SILTP and gLTP. The parameters which give the best average MCA over these 3-folds of the training set were selected as the best parameters. Table 4 shows the ranges of the parameters used for parameter selection.

Comparison of LBP, LTP, SILTP and gLTP

Table 4 reports the experimental results. Using the larger neighborhood ($\{(N = 8, R = 1), (N = 12, R = 2), (N = 16, R = 3)\}$) improves the MCA of all the descriptors regardless of the dataset. It is clear that LBP gives modest performance compared to the best performing descriptor as it is sensitive to noise and illumination changes. As expected SILTP gives better performance than LBP and LTP for CK-dataset images as they are affected by illumination, and LTP gives better performance compared to LBP and SILTP on cell dataset as the images in that dataset are severely affected by noise. Neither LTP nor SILTP gives better performance on either dataset. The proposed gLTP descriptor gives similar performance compared to the best performing SILTP on CK-dataset, and performs significantly better than others on the cell images. Note that the aim of the experiments based on gLTP is not to beat other LBP-based descriptors, but to show that under different conditions gLTP gives stable performance compared to others.

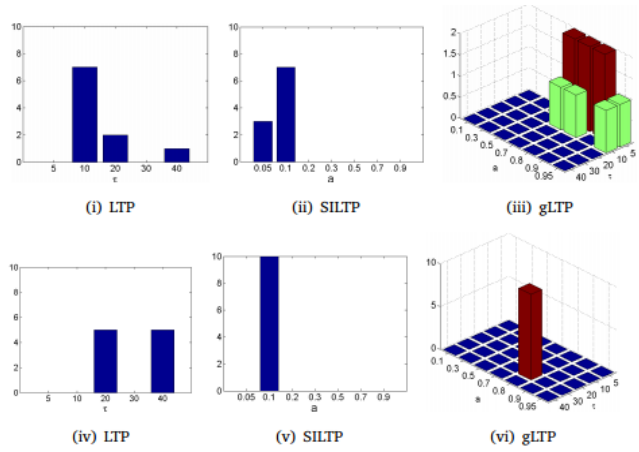


Figure - 9 Histogram of the selected parameters for LTP, SILTP and gLTP

CONCLUSION

This paper presents a generalized version of LBP, LTP and SILTP descriptor called the generalized *LTP* (gLTP). LBP are sensitive to noise as well as illumination changes; LTP are robust to noise but not to illumination changes; SILTP are robust to illumination changes but not to noise. Instead the proposed gLTP are robust to *both* noise *and* illumination changes. I experimentally showed on CK-Dataset and cell images, where the CK- images were taken under different illumination conditions and the images were severely affected by noise, that neither LTP nor SILTP gives better performance on either dataset. On the other hand, the proposed gLTP gives competitive performance compared to

the best performing descriptors on the datasets, confirming that the gLTP is robust to noise and illumination changes.

REFERENCES

[1] M. Brown, G. Hua, and S. Winder, "Discriminative learning of local image descriptors," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 1, pp. 43–57, 2011.

[2] L. Cao, R. Ji, Y. Gao, Y. Yang, and Q. Tian, "Weakly supervised sparse coding with geometric consistency pooling," in IEEE Conference on Computer Vision and

[3] Pattern Recognition, 2012.

[4] T. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "PCANet: A simple deep learning baseline for image classification," IEEE Transactions on Image Processing, vol. 24, pp. 5017–5032, 2014.

[5] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," ACM Transactions on Intelligent Systems and Technology, vol. 2, pp. 27:1–27:27, 2011.

[6] F. Chang, C. Lin, and C. Lu, "Adaptive prototype learning algorithms: Theoretical and experimental studies," Journal of Machine Learning Research, vol. 7, pp. 2125–2148, 2006.

[7] T. Chen, K.-H. Yap, and L.-P. Chau, "From universal bag-of-words to adaptive bag-of-phrases for mobile scene recognition," in International Conference on Image Processing, 2011, pp. 825–828.

[8] E. Ciaccio, G. Bhagat, C. Tennyson, S. Lewis, L. Hernandez, and P. Green, "Quantitative assessment of endoscopic images for degree of villous atrophy in

[9] celiac disease," Journal of Digestive Diseases and Sciences, vol. 56, pp. 805–811, 2011.

[10] A. Coates and A. Y. Ng, "Learning feature representations with k-means." In Neural Networks: Tricks of the Trade (2nd ed.). Springer, 2012, vol. 7700, pp. 561–580.

[11] A. Coates, A. Y. Ng, and H. Lee, "An analysis of single-layer networks in unsupervised feature learning," in Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, 2011.

[12] F. Cohen, Z. Fan, and M. Patel, "Classification of rotated and scaled textured images using gaussian markov random field models," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 13, no. 2, pp. 192–202, 1991.

[13] M. Coimbra, P. Campos, and J. Cunha, "Extracting clinical information from endoscopic capsule exams using MPEG-7 visual descriptors," in The 2nd European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology, 2005, pp. 105–110.

[14] "Topographic segmentation and transit time estimation for endoscopic capsule exams," in IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 2, 2006, pp. II–II.

[15] M. Coimbra and J. Cunha, "MPEG-7 visual descriptors contributions for automated feature extraction in capsule endoscopy," IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 5, pp. 628–637, 2006

[16] L. Cui, C. Hu, Y. Zou, and Meng, "Bleeding detection in wireless capsule endoscopy images by support vector classifier," in *IEEE International Conference on Information and Automation*, 2010. [40] F. DA, M. JT, and e. a. Ben-Menachem T, "Complications of colonoscopy," *Journal of Gastrointestetinal Endoscopy*, vol. 74, pp. 745–752, 2011.

[17] J. P. da Silva Cunha, M. T. Coimbra, P. Campos, and J. M. Soares, "Automated topographic segmentation and transit time estimation in endoscopic capsule exams." *IEEE Transactions on Medical Imaging*, vol. 27, no. 1, pp. 19–27, 2008.