

Crowd Estimation Using Histogram Model Classification Based on Improved Uniform Local Binary Pattern

Seyed Mojtaba Mousavi, Seyed Omid Shahdi, and S. A. R. Abu-Bakar

Abstract—Estimating crowd density may be a good solution for control management, maintaining the crowd safety, or prevention of riot and high risk activities. This paper presents a computational fast and simple method for estimating crowd density based on histogram model classification. The histogram model here is based on the proposed Improved Uniform Local Binary Pattern features. Two main advantages of using this improved version of the local binary pattern are that the pattern features are now intensity invariant as well as rotational invariant. Our proposed method also uses less number of features which makes it faster without sacrificing the overall performance. It has been shown that this method is robust in areas with very low, low, and medium crowd densities. Performance and comparisons with the original local binary pattern method are demonstrated in experimental results.

Index Terms—Crowd estimation, histogram model classification, binary pattern

I. INTRODUCTION

As the number of the world population increases, varieties of challenges are also in the rise. Among others, security is a great concern whenever a large number of people gather in some public premises. Therefore, many public areas such as airports, stadiums, and subways are employing crowd monitoring systems to ensure public safety. These systems normally involve the use of closed circuit television (CCTV) where people and their behavior are monitored continuously by security officers. Because the monitoring task is an arduous task, loss of concentration by these humans is inevitable after some period of time. Moreover the observer's judgment may be influenced by different situations. In addition, it is an uncomfortable task when the officer needs to check several CCTV images concurrently. In such a case, automatic crowd density estimation can be a solution for control management and monitoring the crowds. Consequently, developing a solution for estimating crowd density is an interesting field for researchers[1].

Since the last two decades, different kinds of methods have been used for automatic crowd density estimation. These techniques could be categorized into either pixel based or texture-based methods. The former method uses background removal and then edge detection and the latter uses a texture descriptor for estimating the crowd density.

Marana et al.[2] presented Minkowski Fractal Dimension (MFD) method which used the number of pixels for computing the density of captured image. In their method,

after capturing the image, a threshold version of edge detected image would be generated. By performing the different size of dilation on this binary image and then finding the MFD factor as the single feature for each image, and followed by using the Self Organizing Map (SOM) as a classifier, the density of proposed image would then be estimated. This method does not remove the background and hence requires simple background. Such simplicity causes the system to become less effective in many applications with complicated backgrounds. Although the pixel based technique is simple and fast, it is not reliable when the crowd density is high.

The texture-based method uses crowd images of different densities to present various texture patterns; hence the crowd density can be estimated by texture analysis. One of these methods proposed by Xiaohua et al.[3] is based on wavelet transform and support vector machine (SVM). In their algorithm, the crowd image is first transformed into multi-scale formats using wavelet transform. For extracting the density character vectors, the first-order and second-order statistical features at each scale of the transformed images are used. A classifier based on a SVM is designed to classify the extracted density character vectors into different density levels.

The other texture based method made use of the Advanced Local Binary Pattern (ALBP) as texture descriptor, which is a structural method for automatic surveillance system and could generate a density map as well as a total density estimation of a target area[4]. In addition to introducing the ALBP as a generic texture descriptor, the total density of an image is calculated based on image cells inside it. The image is first divided into cells and then manually labeled them during the training step according to their densities. By having these labels, the histogram for each density could be achieved. During the testing step the density of each cell of input image could be estimated by comparing to the model histogram obtained in the training step. In contrast to the pixel-based techniques, these texture-based methods have been applied successfully for images of high crowd density, and are thus believed to be more robust. The drawback in [4] is manual labeling of the selected frames. This is because different experts may not label the same frame in the same way.

The technique proposed in this paper follows the same footsteps as in [4] which first computes the histogram of an image. Here an Improved Uniform Local Binary Patterns (IULBP) is used to efficiently extract the histogram of each frame. After comparing this histogram with 5 reference histogram models, the crowd density within the frame can be determined based on the following five categories: very low, low, moderate, high, and very high. The remainder of this paper is arranged as follows. In section 2, IULBP is introduced. The proposed system for crowd estimation is

Manuscript received April 12, 2012; revised June 3, 2012

The authors are with Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310, Skudai, Johor, Malaysia (e-mail: mosavi@fkegraduate.utm.my, shahdi@fkegraduate.utm.my, syed@fke.utm.my)

given in section 3. The experimental results are presented in section 4. Finally conclusion and future work are provided in section 5.

II. IMPROVED UNIFORM LBP

Initially the basic of Local Binary Pattern (LBP) is reviewed and then the Uniform Local Binary Pattern (ULBP) is briefly discussed before the IULBP is introduced.

A. LBP

LBP is a texture based method to extract the histogram from the gray scale image. This method is fast and simple and has been successfully applied over a wide range of applications such as face recognition, classification of textures, and image background subtraction[5], [6].

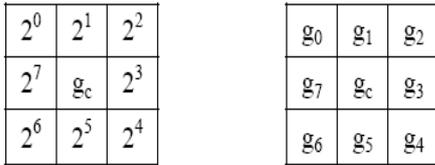


Fig. 1. (a). LBP values in a 3x3 block, (b) counterpart gray values of circularly symmetric neighbors

The first step in Original LBP (OLBP) is to partition the image into N blocks (or sub images) of similar size. In each block, the neighborhood pixels are compared with the value of the central pixel which is chosen as a threshold value. By iterating this process for the whole blocks the histogram will be computed. In this method different size of neighborhoods could be applied to extract the features. The coordinates of row and column, (x_p, y_p) , for each sample (p) is obtained as follow[7], [8]:

$$(x_p, y_p) = (x + R \cos(\frac{2\pi p}{P}), y + R \sin(\frac{2\pi p}{P})) \quad (1)$$

where P represents the total number of sampling points and R is the radius of the neighborhood. Tolerance against illumination changes is the property that makes LBP suitable for real-time applications. Computational simplicity is another equally important advantage of LBP that makes it appropriate for analyzing image in challenging real-time settings, although the significant number of features which are extracted by the OLBP may be useless.

The ULBP produces lower number of features in comparison to the OLBP by utilizing the gray scale invariance and rotation invariance properties. In the following parts, these concepts are explained.

- Gray scale invariance: in order to achieve this goal the following formula is used[9-11]:

$$LBP_{P,R} = \sum_{p=1}^P \text{sign}(g_p - g_c) 2^p \quad (2)$$

$$\text{sign}(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

where sign is a sign operator; g_c and g_p are gray value of the central and circular neighborhood pixels respectively which is demonstrated in Fig 1.

- Rotation invariance: to eliminate the rotational effect a unique identifier could be assigned to each rotation invariant LBP code. This equation is defined by:

$$LBP_{P,R}^{ri} = \min\{ROR(LBP_{P,R}, j) | j = 0, 1, \dots, P - 1\} \quad (3)$$

Where $ROR(x, j)$ produces j times circular bit-wise right shift on the P -bit number x . Rotation invariant LBP include uniform patterns and non-uniform patterns.

In uniform patterns there are at most two transitions between 0 and 1 and in non-uniform patterns there are more than two transitions between them (see Fig.2). Although the ULBP features are less than OLBP, still it is a time consuming process. Moreover by using the raw image as an input to the ULBP, the noise could affect the outcomes[12, 13]. The ULBP histogram is extracted by using (4) and (5) is used for calculating the number of transition in a ULBP code.

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{p=0}^{P-1} \text{sign}(g_p - g_c), & \text{if } U(LBP_{P,R}) \leq 2 \\ P + 1, & \text{otherwise} \end{cases} \quad (4)$$

$$U(LBP_{P,R}) = |\text{sign}(g_{P-1} - g_c) - \text{sign}(g_0 - g_c)| + \sum_{p=1}^{P-1} |\text{sign}(g_p - g_c) - \text{sign}(g_{p-1} - g_c)| \quad (5)$$

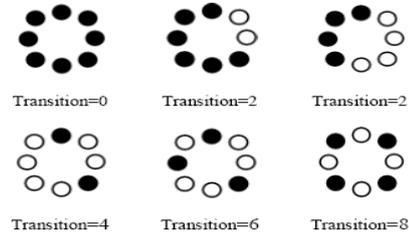


Fig. 2. Uniform and non-uniform patterns are illustrated in the first and second row respectively. black and white circles correspond to bit values of 0 and 1 in the 8-bit output of LBP.

B. IULBP

In the proposed IULBP method there is a modification to reduce both dimensionality and noise in the input frame. This enhancement is achieved by using a blurred image.

High frequency contents of the image are removed by the blurring process. In crowd estimation applications, these details do not contribute to improving the result. By removing these details, the computation time is reduced. In this process the input frame is partitioned into N equal size of blocks (BS). Here the key issue is the total number of samples (P) in each block. Generally, by using more samples the accuracy of the system becomes high as more features can be represented by the histogram which in return helps to improve the results. Equation (6) is the blurring equation:

$$\text{blur}(k, l) = \frac{1}{BS^2} \{ \sum_{a=a_1}^{a_2} \sum_{b=b_1}^{b_2} f(a, b) \} \quad (6)$$

$$a_1 = [BS * (k - 1) + 1], a_2 = BS * k$$

$$b_1 = [BS * (l - 1) + 1], b_2 = BS * l$$

where f is the input frame, a and b are the range of row and column of input frame respectively, l and k are coordinates of row and column of the resulting blurred image.

The fundamental characteristic of the LBP patterns is that most of them are having uniform patterns[7]. Due to this fact, the majority of histogram labels should be allocated to uniform patterns. To accelerate this purpose, $(P + 1)$ uniform binary patterns, which are produced by a circular neighbor set of P sample points, are located into the label bins between 1 and $(P + 1)$ in the histogram, meanwhile the non-uniform patterns are clustered into the histogram label bin $(P + 2)$ as shown in Fig.3.

III. CROWD ESTIMATION SYSTEM

In this paper we propose to estimate the density of the people from the input frame using IULBP. The system consists of two phases, i.e. training and testing, which is discussed separately as follow:

A. Training Phase

The goal for the training phase is to find the model histogram for each of the following category of the crowd densities: Very Low (VL), Low (L), Moderate (M), High (H) and Very High (VH).

During training session the system learns how to estimate the number of people within the input frame according to its histogram. Initially the input frame is given to the IULBP. The output of IULBP is a histogram which is dependent on total number of samples (P) in each block and also the size of block mean. For each category of crowd density, the system is trained to obtain a unique model histogram. This task is done by averaging all of the histograms which belong to the same group of crowd density.

B. Testing Phase

In testing part, the system categorizes the crowd density of input frame into one of the mentioned five categories. This is achieved by computing the distance, D , between the model histogram, H_M , and the test histogram, H_T , as used in [7].

$$D(M, T) = \sum_{b=0}^{P+1} |H_M(b) - H_T(b)| \quad (7)$$

The similarity measure (L) then can be computed by taking the inverse of D .

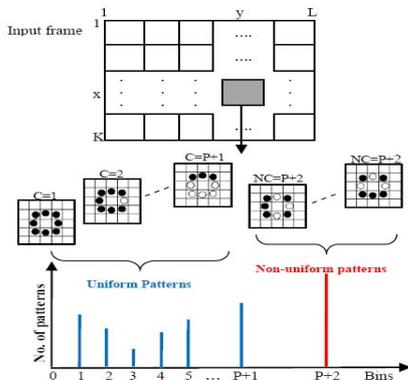


Fig. 3. Process of allocating histogram values, where K and L are number of blocks in row and column of input frame respectively. The C depicts the uniform pattern block correspond to its unique ULBP code. all non-uniform patterns (NC) are assigned to bin number $P+2$.

IV. EXPERIMENTAL RESULTS

In conducting our experiments, we used Intel core2 Duo CPU with 4GB RAM. As for the software, we utilize MATLAB programming environment version R2008a. We used a normal camera to record standard VGA size of 640×480 pixels for streaming video.

A. Database

We used a video sequence collected at a soccer field. This video sequence contained 14000 frames of which 2000 frames were randomly selected such that to contain five categories of crowd densities. From these selected frames, 1000 frames were used for the training and the other 1000

frames were used for the testing. Broadly speaking, each of the five categories contains 200 frames for training and another 200 frames for testing.

In determining crowd densities, we followed the following rules: less than 5 people are categorized as Very Low density, from 5 to 8 people as Low density, between 9 and 12 as Moderate density, between 13 and 16 as High density, and finally 17 or higher as Very High density. Fig.4 shows examples of these categories.

B. Training the System and Obtaining Model Histogram

In order to verify the robustness of our method, we used two different block sizes, and for each size we compute the histogram using both block mean approach (IULBP) and without the block mean approach (ULBP). Hence, during the training sessions we obtained 4 types of model histograms as given below:

- Histogram models using block mean of size 8 and 16 sampling points: Prior to the feature extractor, average of the input frame need to be obtained, according to the defined block size e.g. 5×5 pixels. Furthermore, the IULBP feature extractor uses the $(P, R) = (8, 1)$ or $(16, 2)$ as the neighborhood size.
- Histogram models without using block mean of size 8 and 16 sampling points: The histogram model can be directly obtained by implementing ULBP method using two sampling sizes i.e. $(P, R) = (8, 1)$ or $(16, 2)$.

C. Testing the System

Once the histogram model for each of the five categories has been obtained, the performance of the system can be tested using those 200 testing frames randomly selected. According to (7), the winning category is the one that has the minimum distance or maximum similarity. Fig.5 demonstrates examples to detect the crowd density for each category. These results are obtained using IULBP with 16 sampling points.



Fig. 4. Sample for different kinds of density: (a) Low, (b) Moderate, (c) High, (d) Very high.

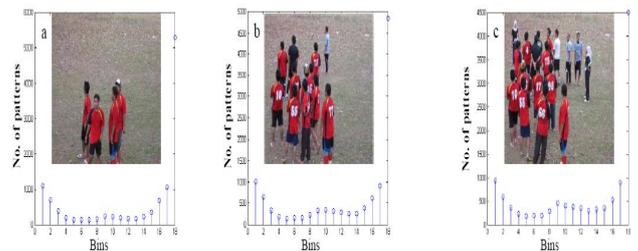


Fig. 5. Crowd density images and their corresponding histograms: (a) Very low, (b) Moderate, (c) Very high.

D. Results

We tabulated our findings between IULBP and ULBP methods for 16 sampling points. Table 1 shows the results of

such findings. From this table it can be found that when the number of people increases, the accuracy of the system decreases. Also among the all mentioned methods, one which uses the block mean and more sampling points have better performance. Table 2 lists the overall average system performance for both 8 and 16 sampling points.

V. CONCLUSION AND FUTURE WORK

In this paper we have demonstrated the performance of the improved uniform local binary pattern (IULBP) method in categorizing crowd density into five categories: very low, low, moderate, high and very high. From the IULBP we first obtained the histogram models that fit into these five categories. Given a testing or target frame, we proceed with computing the IULBP and find its corresponding histogram model. We then compare this target histogram with the five histogram models by computing the distance (D). The one that gives the lowest distance is considered as the winning category.

The advantage of using IULBP is its computational efficiency and also its ability to tolerate the illumination changes. This makes this proposed method appropriate for real-time applications.

The other area of worth undertaking for future work is to combine information coming in from two or more cameras with the IULBP method so that the flow of the crowd can also be estimated.

TABLE I: PERCENTAGE OF RECOGNITION USING 16 SAMPLING POINTS

		VL	L	M	H	VH
VL	Block Mean	95	5	0	0	0
	W/O Block Mean	92	8	0	0	0
L	Block Mean	0	90	10	0	0
	W/O Block Mean	9	86	0	0	5
M	Block Mean	0	2	93	5	0
	W/O Block Mean	0	5	73	22	0
H	Block Mean	0	0	5	85	10
	W/O Block Mean	0	0	32.5	47.5	20
VH	Block Mean	0	0	0	20	80
	W/O Block Mean	0	0	0	25	75

TABLE II: COMPARE OVERALL CORRECT AVERAGE FOR ALL METHODS

IULBP with (P,R) = (16,2)	88.5%
IULBP with (P,R) = (8,1)	84.2%
ULBP with (P,R) = (16,2)	76%
ULBP with (P,R) = (8,1)	64.5%

REFERENCES

[1] H. Rahmalan, M. S. Nixon, and J. N. Carter, "On Crowd Density Estimation for Surveillance. In Crime and Security," *The Institution of Engineering and Technology Conference on*, 2006.
 [2] A. N. Marana, L. D. F. Costa, R. A. lotufo, and S. A. Velastin, *Estimating crowd density with Minkowski fractal dimension*, in

Acoustics, Speech, and Signal Processing, 1999. ICASSP '99. Proceedings. 1999 IEEE International Conference on. 1999.
 [3] L. Xiaohua, S. Lansun, and L. Huanqin, *Estimation of Crowd Density Based on Wavelet and Support Vector Machine*. Transactions of the Institute of Measurement and Control, vol. 28, no. 3, pp. 299-308. 2006.
 [4] M. Wenhua, H. Lei, and L. Changping, "Advanced Local Binary Pattern Descriptors for Crowd Estimation," in *Computational Intelligence and Industrial Application*, 2008. PACIIA '08. Pacific-Asia Workshop on. 2008.
 [5] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Gray Scale and Rotation Invariant Texture Classification with Local Binary Patterns," in *Computer Vision - ECCV 2000*, Springer Berlin / Heidelberg. p. 404-420, 2000.
 [6] T. Mäenpää, "Handbook of Pattern Recognition and Computer Vision," t. edition, Editor. 2004.
 [7] S. W. Lee and S. Z. Li, *Multi-scale Local Binary Pattern Histograms for Face Recognition*, in *Advances in Biometrics*, Springer Berlin / Heidelberg, pp. 809-818, 2007.
 [8] L. Shu, F. Wei, C. S. C. Albert, and Y. D. Yan, "Facial Expression Recognition using Advanced Local Binary Patterns, Tsallis Entropies and Global Appearance Features. in Image Processing," *IEEE International Conference on*, 2006.
 [9] S. Caifeng, G. Shaogang, and P. W. M. Owan, "Robust facial expression recognition using local binary patterns," in *Image Processing, ICIP 2005, IEEE International Conference on*, 2005.
 [10] M. Bingpeng, Z. Wenchao, S. Shiguang, C. Xilin, and G. Wen, "Robust Head Pose Estimation Using LGBP. In Pattern Recognition," *ICPR 2006, 18th International Conference on*, 2006.
 [11] T. Pajdla and J. Matas, *Face Recognition with Local Binary Patterns*, in *Computer Vision - ECCV 2004*, Springer Berlin / Heidelberg. pp. 469-481, 2004.
 [12] L. Shu and A.C.S. Chung, *Texture Classification by using Advanced Local Binary Patterns and Spatial Distribution of Dominant Patterns*. in *Acoustics, Speech and Signal Processing, ICASSP 2007, IEEE International Conference on*, 2007.
 [13] K. D. Su, L. W. Heon, and K. I. So, "Automatic edge detection using 3 ideal binary pixel patterns and fuzzy-based edge thresholding," *Pattern Recogn. Lett*, vol. 25, no. 1, pp. 101-106, 2004.



(Malaysia).

Seyed Mojtaba Mousavi received the B.S. degree in Telecommunication engineering from the Azad University of Najaf-Abad, Iran, in 2003 and the M.S. degree in electronics and telecommunication engineering from the Universiti Teknologi Malaysia (UTM), in 2010. Since 2011, he has been working toward the PhD degree in the computer vision, video & image processing lab, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Johor



Seyed Omid Shahdi received the B.S. degree in electrical engineering from the Azad University of Yazd, Iran, in 2006 and the M.S. degree in electrical engineering from the Azad University of Qazvin, Iran, in 2009. Since 2009, he has been working toward the PhD degree in the computer vision, video & image processing lab, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Johor (Malaysia)



Syed A. R. Abu-Bakar received his BSc (Electrical Engineering) degree from Clarkson University in Potsdam, New York (USA), in 1990, and an MSEE degree from Georgia Tech in 1991 (USA), and the PhD degree from the University of Bradford, England in 1997. In 1992, he joined Universiti Teknologi Malaysia as a lecturer in the Faculty of Electrical Engineering. Currently he is an associate professor in the same faculty. In 2004, he setup the Computer Vision, Video and Image Processing research lab and has become the head since then. His current research interests include image processing with applications in medical imaging, biometrics, agricultural and industrial applications, and computer vision especially in the area of security and surveillance. He has published more than 100 scientific papers both at national and international levels. He is also a senior member of IEEE.