2012 IEEE International Conference on Multimedia and Expo Workshops

# Crowd Density Estimation Based On Local Binary Pattern Co-Occurrence Matrix

Zhe Wang, Hong Liu, Yueliang Qian and Tao Xu
Key Laboratory of Intelligent Information Processing &&
Beijing Key Laboratory of Mobile Computing and Pervasive Device
Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
{wangzhe01, hliu, ylqian, xutao}@ict.ac.cn

***Abstract*—Crowd density estimation is important for intelligent video surveillance. Many methods based on texture features have been proposed to solve this problem. Most of the existing algorithms only estimate crowd density on the whole image while ignore crowd density in local region. In this paper, we propose a novel texture descriptor based on Local Binary Pattern (LBP) Co-occurrence Matrix (LBPCM) for crowd density estimation. LBPCM is constructed from several overlapping cells in an image block, which is going to be classified into different crowd density levels. LBPCM describes both the statistical properties and the spatial information of LBP and thus makes full use of LBP for local texture features. Additionally, we both extract LBPCM on gray and gradient images to improve the performance of crowd density estimation. Finally, the sliding window technique is used to detect the potential crowded area. The experimental results show the proposed method has better performance than other texture based crowd density estimation methods.**

***Keywords - crowd, density estimation, texture, Local Binary Pattern, co-occurrence matrix***

## I. INTRODUCTION

With the growth of population and the worldwide urbanization, crowd phenomenon has become more and more frequent. Crowd density estimation is an important issue in intelligent video surveillance, which becomes an important research approach of computer vision in recent years. Polus et al.[1] first introduce the level of services for a pedestrian flow, which is widely adopted. Based on this idea, crowd density can be defined as: free flow, restricted flow, dense flow and jammed flow. In real world surveillance application, different crowd density levels maybe need different attention.

In recent years, many crowd density analysis methods have been proposed. Zhan et al. [2] make a survey on crowd analysis. In this survey, crowd density estimation methods can be divided into three categories. Firstly, crowd analysis based on background removal techniques [3][4][5]. Secondly, crowd analysis based on image processing and pattern recognition techniques [6][7][8]. In this category, texture features are widely used. Thirdly, crowd analysis based on information fusion [9][10].

Marana et al.[6] consider that high density crowd has fine patterns of texture, while images of low density have coarse patterns of texture. Based on this assumption, many image texture features have been used in crowd density estimation. Some crowd density estimation algorithms, like GLCM [6] and GOCM [11], extract features from the whole image and give an overall density level estimation for the whole image. However, there are two main shortages of this kind of method. One is that pedestrian may only appear in a certain area of the image, such as pedestrian would be only in the walkway. The other is that in real-world applications the density level of a specific area is crucial. For instance, in the stadium, the density level of the exit is more important than that of the spectator seats.

Local Binary Pattern (LBP) has been widely used and proved to be an effective texture descriptor for texture analysis [8][12] and it has also been used in crowd density estimation. Ma et al. [8] calculate LBP code in block mean domain and use Dual-Histogram LBP to reduce dimensions of the feature. Yang et al. [12] propose a sparse spatiotemporal LBP algorithm and utilize its statistical property to describe the crowd feature. These methods use histogram of LBP in crowd density estimation and just ignore the spatial property of LBP, which would improve the classification accuracy.

In this paper, we propose a novel strategy to use LBP for crowd density estimation. Firstly, we present a novel LBP Co-occurrence Matrix (LBPCM) based algorithms to extract texture features. It contains the spatial information of LBP while other histogram based LBP descriptors often neglect. Moreover, overlapping cells are used to extract texture feature vector, which can code more local information. Secondly, we extract LBPCM on gray image and gradient image respectively and then concatenate them to construct one feature vector. Experimental results show the proposed method can achieve significant improvement than other existing texture features.

## II. OVERVIEW OF OUR APPROACH

We propose a sliding-window based framework to classify and locate crowd area as Fig.1 shows. Firstly, we cut image blocks from the original images which varies in crowd density levels, background and illumination condition. These blocks are annotated with the corresponding crowd density level according to [1]. Secondly we divide each block into several overlapping cells and then texture features are calculated from each cell. We concatenate these features to construct one feature vector, which is the texture descriptor of an image block. After calculating all the feature vectors of the image blocks, we use SVM to train classification modes for each crowd density level. In the detecting stage, we use a

978-0-7695-4729-9/12 $26.00 © 2012 IEEE
DOI 10.1109/ICMEW.2012.71

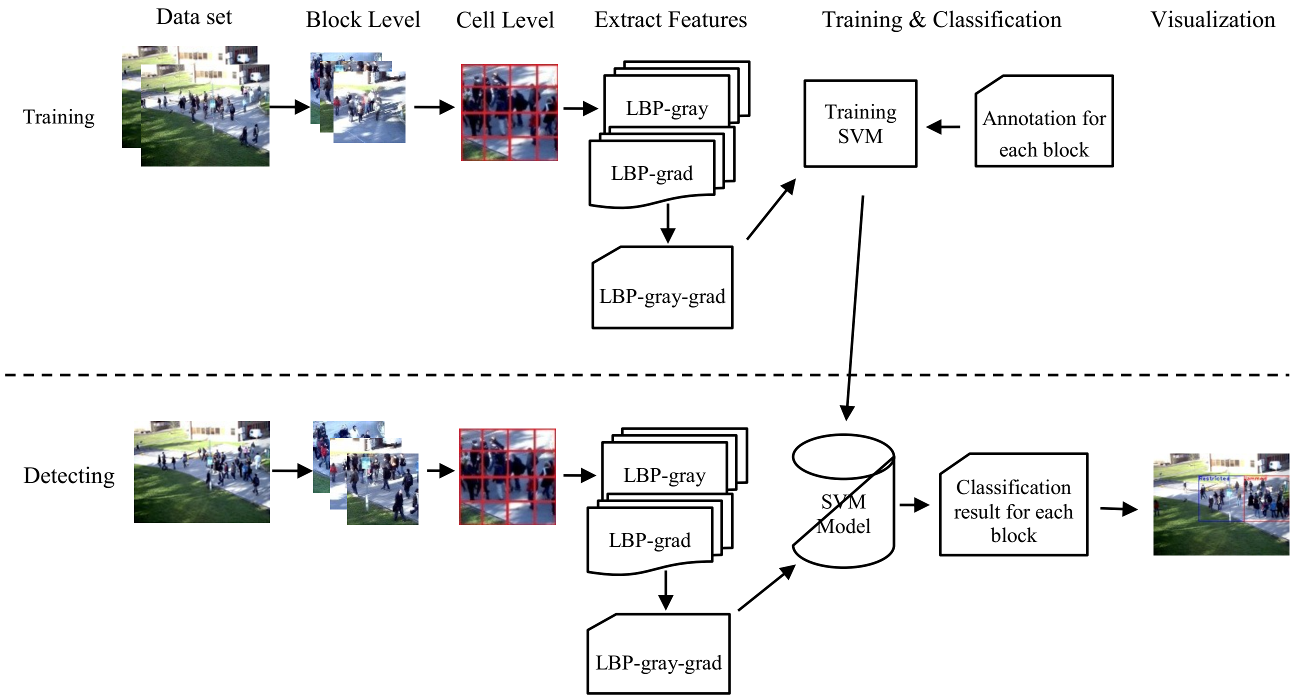



Figure 1. Flow chart of the LBPCM based crowd density analysis system.

sliding window to scan the given image and give different remarks for each detection window with different crowd density levels. Fig. 1 shows the flow chart of our system.

## III. PROBLEM FORMULATION

The problem of crowd density estimation can be defined as following. Let $X = \{(x_i, c_i)_{i=1,\ldots,N}\}$ denote the training set, where $x_i \in X$ is a training image block and $c_i \in \{0, 1, 2, 3\}$ denotes the crowd density level: free flow, restricted flow, dense flow and jammed flow. The texture features are calculated for each $x_i \in X$ and finally we have the feature set $F = \{(f_i, c_i)_{i=1,\ldots,N}\}$., where $f_i$ represents the texture feature for image block $x_i$. Then the feature set $F$ is trained by SVM to get a SVM model $M$. Next, each testing sample $y_i$ of the testing set $Y$ is calculated to get its texture feature descriptor. Then model $M$ classify $y_i$ into a density level $c_i \in \{0, 1, 2, 3\}$. Continuing with this procedure, each testing sample can be classified.

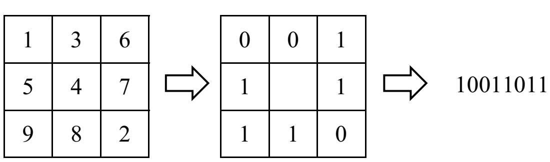



Figure 2. An example of LBP.

## IV. LBP CO-OCCURRENCE MATRIX (LBPCM)

Marana et al. [6] consider that high density crowd has fine patterns of texture, while images of low density have coarse patterns of texture. Based on this assumption, many image texture features have been used in crowd density estimation. Among these features, LBP has been widely used and proved to be a powerful tool for different kinds of problems such as face recognition, human detection and texture analysis [8][12]. In this paper, we propose a novel strategy to use LBP for crowd density estimation.

### *A. Local Binary Pattern Image*

LBP is proposed by Ojalaet al. [13] in 1994. And it has been found to be a valuable texture feature. As Fig.2 shows, LBP is calculated one pixel by another. For each pixel, it is compared to its neighbors (number of neighbors can be 8, 12 or even more). Then follow the neighbors clockwise or counter-clockwise. If its value is bigger than the center pixel, write "1". Otherwise, write "0". When traversed all the neighbors, an 8-digit binary number can be obtained. Usually it is converted to decimal. In most cases, the histogram is calculated by LBP over the cells in the detection window and is used as classification features.

In our work, LBP image is constructed for further processing. We calculate LBP for each pixel with distance as 1 pixel and consider all the 8 neighbors. Then each pixel, except the boundary of the input image block, gets a LBP value. The LBP value of a pixel is exactly an integer between

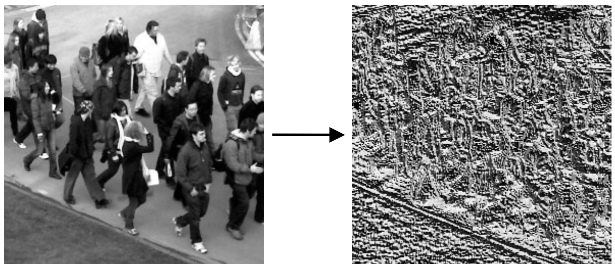


Figure 3. LBP image. Left: a gray scale image. Right: the corresponding LBP image.

0 and 255, which makes the LBP map of the whole image block can be processed as a gray scale image. Fig. 3 shows a gray image block and its LBP texture image.

LBP code can well describe the local texture pattern. While in a large area of an image, the statistical and spatial information of LBP may not strong enough for detection and classification. So we divide each block into several overlapping cells and extract LBPCM features of each cell. Then all the cell features are concatenated to construct a feature vector that describes the texture of the image block. In order to know what the best cell size is, we choose different cell size and different step to evaluate the performance in our experiment.

### *B. Calculate Co-Occurrence Matrix on LBP Image*

Co-occurrence matrix is proposed by Haralick et al. [14] and Marana et al.[6] use it to estimate crowd density first. Typically, co-occurrence matrix is calculated on gray scale image, known as Gray Level Co-occurrence Matrix (GLCM) which is widely used for crowd density estimation. However we find that the LBP value of a pixel is just between 0 and 255 which can be directly used as gray image in GLCM. Co-occurrence matrix is a statistical method based on the estimation of second-order joint conditional probability density functions, $f(i, j \mid d, \theta)$.Each $f(i, j \mid d, \theta)$ represents the probability that a pair of grey levels $(i, j)$ occurs at the distance of $d$ along the direction $\theta$ in an image. The original co-occurrence matrix is difficult to be directly used for classification because it contains a large amount of information. So, we adopt four widely used descriptors for texture measurements [14]: energy, contrast, homogeneity and entropy, which are defined as follows:

1) Energy

$$s_{energy}(d,\theta) = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} \mathrm{f}(\mathrm{i},\mathrm{j} \mid \mathrm{d},\theta)^2 \tag{1}$$

2) Contrast:

$$s_{contrast}(d,\theta) = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} (i-j)^2 \mathrm{f}(\mathrm{i},\mathrm{j} \mid \mathrm{d},\theta) \tag{2}$$

3) Homogeneity:

$$s_{\mathrm{homogeneity}}(d,\theta) = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} \frac{\mathrm{f}(\mathrm{i},\mathrm{j} \mid \mathrm{d},\theta)}{1+(i-j)^2} \tag{3}$$

4) Entropy:

$$s_{\mathrm{entropy}}(d,\theta) = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} \mathrm{f}(\mathrm{i},\mathrm{j} \mid \mathrm{d},\theta) \tag{4}$$

We calculate four sets of co-occurrence features with parameters $d$ as 1 and $\theta$ as 0°, 45°, 90° and 135° respectively. All the properties are concatenate together.

It is obvious that the histogram of LBP only represents the distribution of texture features in local area while at the same time lost the spatial interactions between pixels. However co-occurrence matrix has the ability to extract both statistical and spatial information. So we extract texture measurements of co-occurrence matrix based on LBP image of each image block as texture descriptors.

### *C. Combine Gray and Gradient Features*

Gradient map of an image contains much information of edge features. We know that different objects have different edge distributions, which may present unique edge texture between each other. Considering pedestrian, dense crowd has more edges of human, which presents fine patterns of texture while sparse crowd has fewer edges of human, which presents coarse patterns of texture. Thus we also calculate LBPCM on gradient image (LBPCM-grad) for classification, as shown in Fig. 4. As gray image and gradient map can complement with each other, LBPCM-gray and LBPCM-grad are concatenated to construct one texture feature vector (LBPCM-gray-grad). The

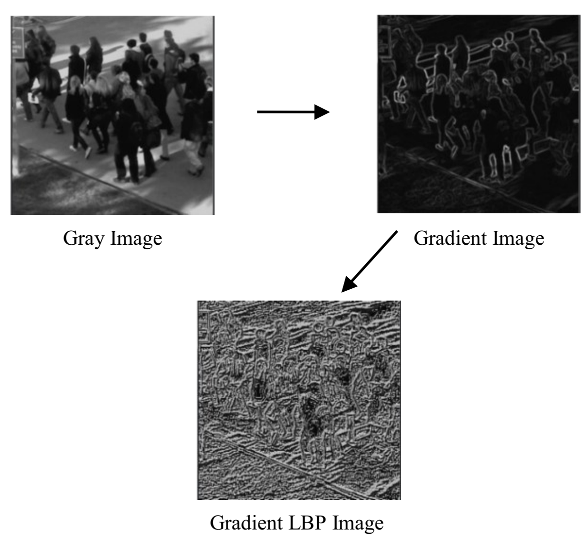



Figure 4. Gradient LBP image. Top-Left: a gray scale image. Top-Right: the corresponding gradient image. Bottom: LBP image of the gradient image.

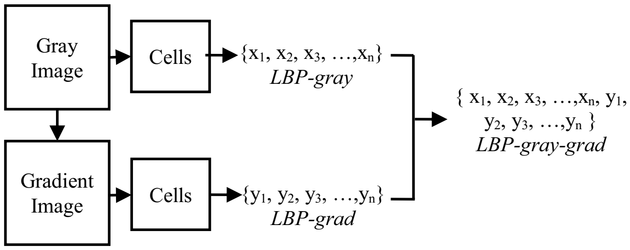



Figure 5. Constructing LBP-gray-grad.

constructing procedure is shown in Fig. 5.

## V. EXPERIMENTAL RESULTS

### A. Dataset

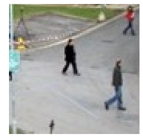
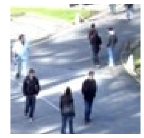
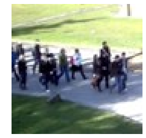
In this paper, we perform experiments on the dataset of PETS 2009 [15] with the resolution of 768×567 pixels. To evaluate the performance of our proposed method, we manually cut image blocks with the size of 256×256 pixels from the original image. All the blocks are divided into four crowd density levels according to the congesting degree of the crowds, which are defined as free flow, restricted flow, dense flow and jammed flow [1]. In addition, all the color images are converted to gray scale before features extraction. We finally get 200 image blocks for each density level, totally 800 blocks. Among them, 100 blocks of each density level are used as training set and the remaining 100 blocks are used as testing set. So we obtain 400 image blocks for training and the other 400 for testing. By split the image blocks into training set and testing set, quantitative analysis and comparison with other algorithms can be conducted. Table 1 gives some image block samples of each density level from dataset of PETS 2009.

### B. Performance Evaluation

#### 1) Evaluation on Different Cell Size

We conduct an experiment to evaluate how cell size can influence the classification accuracy. The results are shown in Fig. 6. As the block is a square, we just use square cells within each block and the horizontal and vertical steps are also the same. Firstly, we look at different cell size with the same step at 32 pixels and compare them together with the accuracy when using features extracting from the original 256×256 pixels block. The highest accuracy is achieved when the cell size is 128×128 pixels, which is half the length of an image block. When the cell size is getting bigger than half of the length of the image, the accuracy is falling lower. When the cell size is smaller than 128×128 pixels, the accuracy also gets lower: the smaller the cell size, the lower the accuracy. This maybe when the cell size is too small, LBPCM contains too little information that can sufficiently describe the local texture. And when the cell size is too large, LBPCM cannot extract the local features. Secondly, we look at different cell steps of the same cell size. For example, when the cell size is 128×128 pixels, the

Table 1. Samples of each density level of PETS Dataset

| Free flow | Restricted flow | Dense flow | Jammed flow |
|---|---|---|---|
|  |  |  |  |

estimation accuracy with step of 32 pixels is higher than that with step of 64 pixels. And the accuracy is lowest when the step is 128 pixels. We can find that the smaller the step the higher the accuracy. All the experimental results demonstrated this. It is logical because when the step is small, more local features can be extracted to describe the texture of the image block. In the following experiments, we let the cell size be 128×128 pixels and with step at 64 pixels to achieve both the classification accuracy and the computational efficiency. All the color images in our dataset are converted into gray images before further processing.

#### 2) Evaluation on Gradient Image

Many texture features directly extract descriptors from gray images because gray images contain rich texture information. LBPCM is also a texture feature, so it is rational that we calculate LBPCM on gray images. We also know that texture has an inherent relationship with the edges because fine patterns of texture tend to have more edges and coarse patterns of texture tend to have fewer edges. Since gradient image contains much information about edges, we consider it is reasonable and valuable to calculate LBPCM on gradient images. So we conduct an experiment to verify that. We both calculate LBPCM features on gray images and the corresponding gradient images respectively. So we have two kinds of texture features: LBPCM-gray and

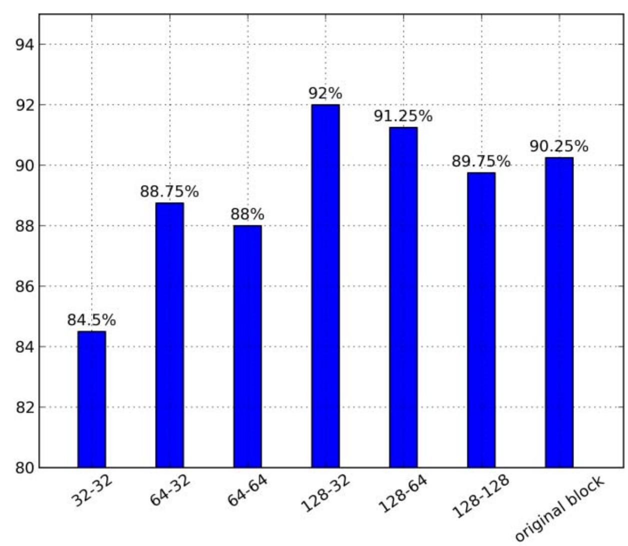



Figure 6. Classification accuracy of extracting LBPCM features from different cell sizes and different steps. "*a-b*" in the horizontal axis denotes cell size is $a \times a$ pixels and cell sliding step is $b$ pixels.

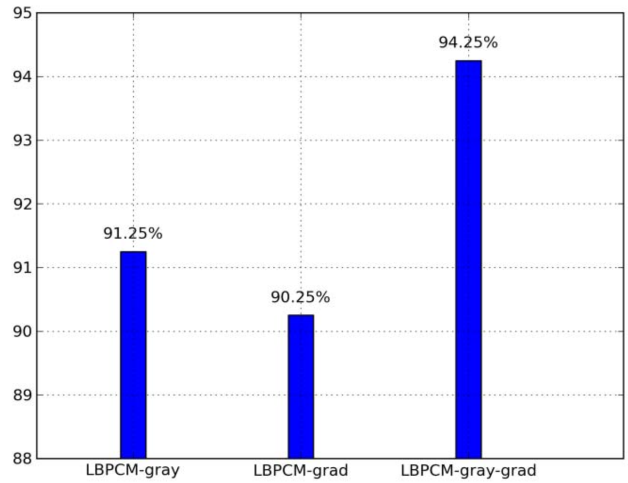



Figure 7. Classification accuracy when using LBPCM-gray, LBPCM-grad and LBPCM-gray-grad.features.

LBPCM-grad. Then we concatenate LBPCM-gray with LBPCM-grad to generate a combo texture feature: LBPC-gray-grad. Now we have constructed three kinds of LBPCM-base texture features, and then these features are used to train SVM classifiers respectively. The classification accuracy on testing set of the three features is shown in Fig. 7. All the features are extracted from 256x256 pixels blocks and the cell size of each block is 128×128 pixels with step at 64 pixels.

From Fig. 7 we can see that LBPCM-gray and LBPCM-grad can well be used for crowd density estimation. This means gradient images also contains valuable textures features as gray images. Furthermore, when LBPCM-gray-grad is used for classification, the accuracy is significantly improved than only LBPCM-gray or LBPCM-grad is used. This experimental result indicates that the combination of gray and gradient features can complement to each other to achieve better description of the texture feature for crowd density estimation.

*3) Comparison with Other Algorithms*

We implement several crowd density estimation algorithms and use them to estimate crowd density level on our data set described in part A of section Ⅴ. The algorithms we used are gray level co-occurrence matrix (GLCM), local binary pattern (LBP) and gradient orientation co-occurrence matrix (GOCM). GLCM is proposed by Haralick et al. [14] and Marana et al.[6]use it to estimate crowd density first. LBP has been introduced in part A of section Ⅳ. In the comparison, the LBP histograms are calculated for each image block using the same cell size and step as LBPCM. The difference is that the LBP descriptor is just the histogram of LBP patterns of a cell. Then each LBP histograms within a block is concatenated to construct one LBP feature vector. Ma et al. [11] described a kind of texture based on gradient map called the gradient orientation co-occurrence matrix (GOCM).The gradient magnitude and orientation are both used in GOCM. The GOCM is a 9×9 matrix, and the calculation algorithm can be found in [11]. All these features together with LBPCM-gray-grad are extracted from the dataset used in our experiment. Then SVM is used to train these features and to classify crowd density levels. Fig. 8 shows the experimental results.

As the graph shows, all the features perform well on classifying crowd density levels. Our feature offers significant higher accuracy than any other features. The LBP histogram can well describe the local texture of an image while the LBPCM we proposed utilizes the spatial information of LBP codes and thus enhances the original LBP to be more descriptive to local area. The combination of LBPCM on gray and gradient features also largely improved the accuracy, which is analyzed in former section. Cell based feature extraction procedure also makes the LBPCM more robust and let LBPCM fully utilize the texture information of the image.

### *C. Crowd Density Estimation on the Whole Image*

In some video surveillance applications, crowd analysis not only needs to classify the crowd density level of current scene but also needs to locate the crowded area. Some applications need to estimate the crowd density at a particular location in the whole image. Our LBPCM based strategy is suitable for this kind of applications and shows well performance. The sliding window technique is adopted in our work. The window size is the same as the image blocks we extracted from the original data set. The sliding step is a parameter, which can be different in different use case. In our example, value of the sliding step is half or one fourth of the length of the sliding window. In this paper, the size of the image block in our experiment is 256×256 pixels and the step of sliding window is set to 128 pixels. As the area of free flow usually does not need attention, we just use our methods to classify and locate restricted, dense and jammed areas in the walkway, which is in the middle of the

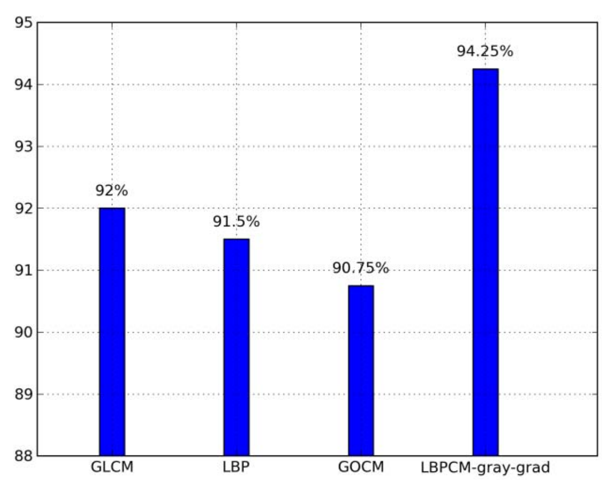



Figure 8. Classification accuracy of different texture features.

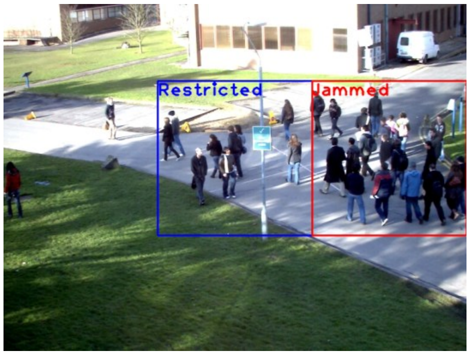



Figure 9. Sample of detection and classification result. Jammed areas are marked with red bounding box.

image in our dataset. In order to handle overlapping detected areas, we label all the jammed areas and merge the overlapping areas. Then we label the dense areas which do not overlap with the already labeled jammed areas, and overlapping dense areas are merged. The restricted areas are labeled in the same way. A sample of the detection result is shown in Fig. 9.

## VI. CONCLUSION

Crowd density estimation is an important issue in real world applications, especially for public security and management. The crowd usually presents some kind of texture features, so pattern recognition methods base on texture have been widely used for crowd density estimation. Among the various texture features, LBP has been found to be a very well descriptor to local texture patterns. It is widely used in computer vision technology based video surveillance applications. However, texture feature by LBP histogram loss the spatial information of LBP. In this paper, we proposed a novel strategy to use LBP features. We calculate the co-occurrence matrix on the LBP map instead of using histogram. Then the cell based method is used to construct the LBPCM feature vector. In our work, texture features on gradient images and gray images are combined together to achieve a better performance. Experimental results have proved the effectiveness of the proposed method. At last, we give a demonstration for local crowd density estimation and location using LBPCM in video surveillance applications.

In future work, we are going to improve the computational efficiency of LBPCM and make it more robust to scale variance.

## ACKNOWLEDGMENT

The research work is supported by the National Nature Science Foundation of China No. 60802067.

## REFERENCES

[1] A. Polus, J. Schofer and A. Ushpiz, "Pedestrian Flow and Level of Service," J. Transportation Eng. 109(1), 46-56, 1983.

[2] Beibei Zhan, Dorothy N. Monekosso, Paolo Remagnino, Sergio A. Velastin and Li-QunXu, "Crowd analysis: a survey," Machine Vision and Applications, vol. 19, pp. 345-357, Oct. 2008.

[3] Jia Hong Yin, Sergio A. Velastin, Anthony C. Davies, "Image processing techniques for crowd density estimation using a reference image," vol. 1035, pp. 489-498, 1996.

[4] R. Ma, L. Li, W. Huang and Q. Tian, "On pixel count based crowd density estimation for visual surveillance," IEEE Conference Cybernetics and Intelligent Systems, vol.1, pp. 170- 173, 1-3 Dec. 2004.

[5] Lan Dong, Parameswaran V., Ramesh V. and Zoghlami I., "Fast Crowd Segmentation Using Shape Indexing," Computer Vision,IEEE 11th International Conference on, pp.1-8, 14-21, Oct. 2007

[6] A. N. Marana, S. A. VelaStin, L. F. Costa and R. A. Lotufo., "Estimation of crowd density using image processing," Proc. IEE Colloquium Image Processing for Security Applications, 11:1-8, 1997.

[7] Tommy W.S. Chow, Siu-Yeung Cho and Chi-Tat Leung, "A Neural-Based Crowd Estimation by Hybrid Global Learning Algorithm," IEEE transation on systems, man, and cybernetics, 29(4), 1999.

[8] Wenhua Ma, Lei Huang and Changping Liu, "Advanced Local Binary Pattern Descriptors for Crowd Estimation," Computational Intelligence and Industrial Application, 2008. PACIIA '08. Pacific-Asia Workshop on , vol.2, no., pp.958-962, 19-20 Dec. 2008.

[9] D.B. Yang, H.H. Gonzalez-Banos, L.J. Guibas, "Counting people in crowds with a real-time network of simple image sensors," Computer Vision, Ninth IEEE International Conference on , pp.122-129 vol.1, 13-16, Oct. 2003.

[10] S.A. Velastin, J.H. Yin, A.C. Davies, M.A. Vicencio-Silva, R.E. Allsop and A. Penn, "Automated measurement of crowd density and motion using image processing," Road Traffic Monitoring and Control, Seventh International Conference on , pp.127-132, 26-28, Apr 1994.

[11] Wenhua Ma, Lei Huang and Changping Liu, "Crowd density analysis using co-occurrence texture features," Computer Sciences and Convergence Information Technology (ICCIT), 2010 5th International Conference on , pp.170-175, Nov. 30 2010-Dec. 2 2010

[12] Hua Yang, Hang Su, ShibaoZheng, Sha Wei and Yawen Fan, "The large-scale crowd density estimation based on sparse spatiotemporal local binary pattern," Multimedia and Expo (ICME), 2011 IEEE International Conference on , vol., no., pp.1-6, 11-15 July 2011

[13] T. Ojala, M. Pietikäinen and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions," Proceedings of the 12th IAPR International Conference on Pattern Recognition, vol. 1, pp. 582 – 585, 1994.

[14] R.M. Haralick, "Statistical and Structural Approaches to Texture," Proceedings of the IEEE, vol.67, pp. 786-804, 1979.

[15] http://www.cvg.rdg.ac.uk/PETS2009/a.html