ICIC Express Letters
Part B: Applications
Volume **11**, Number **5**, May **2020**
ICIC International ©2020 ISSN 2185-2766
pp. 479–485

CROWD DENSITY ESTIMATION USING TAYLOR EXPANSION AND LOCAL BINARY COUNT DESCRIPTOR

Chih-Chin Lai and Hsien-Chun Chiu

Department of Electrical Engineering
National University of Kaohsiung
700, Kaohsiung University Rd., Nanzih District, Kaohsiung 811, Taiwan
cclai@nuk.edu.tw; peanut5638@gmail.com

Received October 2019; accepted January 2020

Abstract. *With the steady population growth, the crowd phenomenon in public areas is becoming more and more frequent. In the past decades, many deadly accidents have been caused by dense human crowds. Therefore, crowd density estimation from images or videos plays an essential role for crowd monitoring and safety control. In this paper, we propose a crowd density estimation method based on the Taylor expansion and the local binary count descriptor. The Taylor expansion is first applied to the crowd image, and then the most discriminative features from the transformed image are extracted by using the block-based local binary count pattern. Finally, these features are classified by the support vector machine. Experiments on the PETS 2009 dataset are provided to demonstrate the feasibility of the proposed approach.*
Keywords: Crowd density estimation, Taylor expansion, Local binary count, Support vector machine

1. **Introduction.** With the growth of population, large-scale gatherings of people can be observed in public places such as airports, subway stations, and shopping malls. Crowds in an area may grow unexpectedly; therefore, a large number of people can be viewed as a potentially dangerous phenomenon. To detect potential risk and to prevent mortal accidents, crowd monitoring is a very important task in the security community for emergency and safety control. Crowd density estimation has become a hot research topic in the intelligent video surveillance area. It is an effective way for crowd monitoring and management. A lot of work has been devoted to studying crowd density estimation in still images [1, 2].

Generally, like most pattern recognition problems, the feature extraction in the crowd density estimation problem is crucial to the whole classification process. In order to estimate the crowd density, crowd features of different density levels should be identified and extracted, and then an appropriate and efficient classification model is used to discriminate the crowd density based on these features. Texture is an important image feature that has been used for characterization of images. Images of crowds with different densities present distinct texture patterns. According to this assumption, Marana *et al.* [3] presented a technique based on the differences of texture patterns of the crowd images to estimate the number of people. To accurately model texture in the crowd image, Wu *et al.* [4] presented a perspective projection approach to generate a series of multi-resolution density cells and to calculate the density distribution of the crowded image by texture analysis and learning technique. Ma *et al.* [5] proposed a patch-based density analysis framework for crowd estimation. For each patch, a novel texture descriptor based on co-occurrence of gradient patterns is extracted and used for classification. Fradi *et al.* [6]

DOI: 10.24507/icicelb.11.05.479

proposed an approach for crowd density estimation at the patch level, and they also proposed to learn a discriminant subspace of the high-dimensional local binary pattern (LBP) raw feature vector where samples of different crowd density are optimally separated. Pai *et al.* [7] proposed a texture feature-based approach for the estimation of crowd density where two texture features namely uniform LBP and Gabor filter are used. The two feature sets are computed separately over different crowd images and concatenated before being given to a classifier for classification.

To estimate the crowd density in an image can be viewed as a texture classification problem. In addition to extracting the discriminative textural features, the selection of a good classifier is also crucial for texture classification. From Bayes classifiers to neural networks, there are many possible choices for an appropriate classifier. Among these, the support vector machine (SVM) [8, 9, 10, 11] would appear to be a good candidate because of its ability to generalize in high-dimensional spaces, such as spaces spanned by texture patterns. Except for the aforementioned texture-based approaches, some techniques for crowd density estimation using SVM have been proposed [12, 13, 14].

Due to its discriminative power and low computational complexity, the LBP [15] and its variants have become a very popular texture descriptor used for classification in different applications of computer vision, image analysis, and pattern recognition [16, 17, 18, 19, 20]. A novel texture descriptor named as local binary count (LBC) [21] is a variant of LBP. It ignores the local binary structure of LBP by only counting the number of value 1's in the pattern. Although they dramatically simplify the geometric structure, both LBC and LBC-like features have been used with success for texture classification. In this paper, we adopt the Taylor expansion to transform a crowd image, and then the block-based LBC descriptor is applied to extracting and encoding the texture features from the transformed image. We use the SVM to classify the texture features for crowd density estimation. Experimental results indicate that the proposed approach can produce a higher classification rate.

The remainder of this paper is organized as follows. In Section 2, the process of applying the Taylor expansion to an image and the basic concept of LBC are introduced, and then the proposed approach is described. The experimental results and discussions are given in Section 3. The conclusions are summarized in Section 4.

2. Methodology.

2.1. Taylor pixel feature.

In some applications of computer vision, we can obtain more useful information from an image in a different representation by mapping the original data set into new data points. In addition to the well-known transformation methods, such as Fourier transform, discrete cosine transform, and discrete wavelet transform, the Taylor expansion has been regarded as an extremely powerful tool for studying the image processing algorithms [22, 23]. To extract more discriminative facial expression features from the image sequence, the Taylor expansion is applied to every pixel in the facial image and then the Taylor pixel feature can be obtained for facial expression classification [24].

The Taylor expansion is the standard technique used to obtain a linear or a quadratic approximation of a function of one variable. Let a pixel at a certain location be denoted as $c = (x_c, y_c)$, and $f_n(c)$ be the nth-order Taylor pixel feature of the pixel c. According to the definition of the Taylor expansion theorem, $f_n(c)$ can be approximately defined as:

$$f_n(c) = \sum_{k=0}^{n-1} \frac{f^{(k)}(\alpha)(g_c - \alpha)^k}{k!} + \frac{f^{(n)}(\alpha)(g_c - \alpha)}{n!}, \tag{1}$$

where g_c is the gray value of the central pixel c. Therefore, the first-order Taylor pixel feature of the pixel c can be expressed as follows:

$$f_1(c) = \frac{f^{(0)}(\alpha)(g_c - \alpha)^0}{0!} + \frac{f^{(1)}(\alpha)(g_c - \alpha)}{1!}, \tag{2}$$

$$f^{(1)}(\alpha) = \begin{cases} 1, & g_c \geq \alpha \\ -1, & g_c < \alpha \end{cases}, \tag{3}$$

where $f^{(0)}(\alpha)$ denotes the average gray value of all the pixels in a 3×3 image region, while α is the average gray value of the 3×3 neighborhoods surrounding the central pixel c.

2.2. **Local binary count.** The basic idea behind LBP is that an image consists of micropatterns. The LBP histogram contains information about the distribution of local micropatterns over the whole image. However, two drawbacks of LBP are 1) it is not very robust against local changes in the texture, and 2) it may not work properly for noisy images or on flat image areas of constant gray level [16]. In order to overcome the aforementioned shortcomings of the existing LBP operator, a novel feature descriptor named LBC is proposed [21]. The LBC can extract the local binary grayscale difference information, and totally abandon the local binary structural information.

The LBC can be viewed as a variant of LBP for texture classification. The basic concept behind LBC and LBP is the same in thresholding the local neighboring pixels, but it differs in the encoding phase by counting the number of value 1's in the local neighbor set of LBC rather than applying binary encoding in the LBP. As a result, the LBC can be defined as follows:

$$LBC_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c), \tag{4}$$

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}, \tag{5}$$

where g_c is the gray value of the central pixel c and g_p is the gray value of the neighbor pixel on a circle of radius R, and P is the total number of the neighbors. Figure 1 gives a comparison of the coding schemes of LBP and LBC in a 3×3 pixel block.

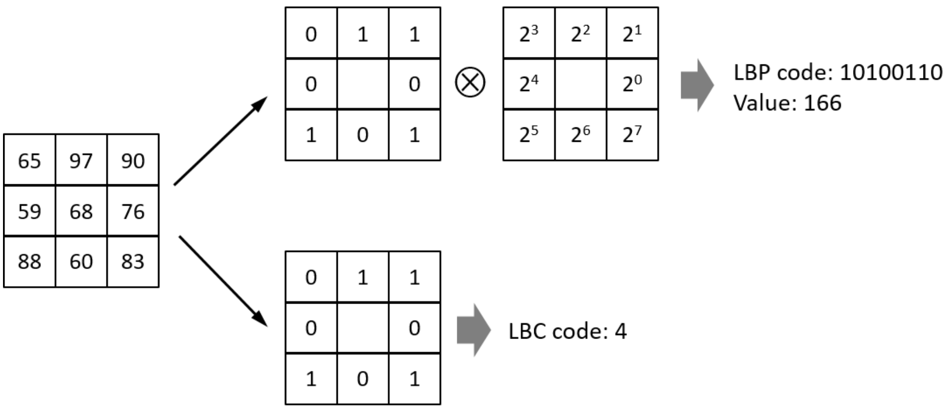

FIGURE 1. Illustration of the coding schemes of LBP and LBC

2.3. **The proposed approach.** In order to retrieve the density information of a crowd image, we develop a two-stage approach to extract the discriminative texture features. First, the Taylor expansion is applied to the crowd image and then the Taylor pixel feature image is obtained. Second, to consider the grayscale difference information with the spatial relationship of the image content, the transformed image is partitioned into small non-overlapping regions $\{R_1, R_2, \ldots, R_M\}$, where each region has the same size commonly. After that, we apply the LBC operator on every pixel in every region. The

LBC histogram for every region is calculated, and then all histograms are concatenated into one feature vector that represents the image. The concatenated feature vector which implicitly embeds the local information provides a better crowd feature representation and describes the image content accurately.

After feature extraction, the next task is to classify the different texture patterns into distinct defined classes with a proper classifier. The SVM is a very powerful and versatile machine learning model [25]. It can be used for data classification or to solve the nonlinear regression problem. In our approach, to classify the crowd density with different texture pattern is performed by adopting an SVM.

SVM has been shown to be very effective because it has the ability to find the optimal separating hyperplane that gives the maximum margin between the positive and negative samples. Given a training set of labeled samples $\{(x_i, y_i),\ i = 1, \ldots, h\}$, where $x_i \in \mathcal{R}^n$ and $y_i \in \{+1, -1\}$, a new test data x is classified by:

$$f(x) = \operatorname{sign}\left(\sum_{i=1}^{h} \varphi_i y_i K(x_i, x) + c\right), \tag{6}$$

where φ_i are Lagrange multipliers of the dual optimization problem, c is a bias or threshold parameter, and $K(\cdot,\cdot)$ is a kernel function. In our work, we used the GPU-based LIBSVM [26] in all experiments.

3. **Experimental Results.** To show the effectiveness of the proposed approach, some experiments on the PETS 2009 [27] are presented. As a public and comprehensive benchmark dataset, the PETS 2009 contains multi-view sequences to show different crowd scenarios in an outdoor environment. These scenarios include: (1) Dataset S1 for person count and density estimation, (2) Dataset 2 for tracking of individual(s) within a crowd, and (3) Dataset 3 for detection of flow and crowd events. In this paper, we mainly conducted experiments on some frames from S1 and S2 for the evaluation of crowd density estimation.

Initially, we manually divide the frames into blocks of size 226×226 which corresponds to an area approximately equal to 13 m^2 based on the descriptions of Fradi *et al.* [6]. Table 1 shows the different crowd density levels according to the range of people in 13 m^2, and some image block samples from the PETS 2009 dataset are shown in Figure 2. Afterward, we labeled these image blocks according to the congesting degrees of the crowd defined in Table 1. Because the number of people in the image blocks could not reach the jammed flow level, only four levels are used for experiments.

Table 1. Different level of crowd density

Crowd density levels	Number of people (for 13 m^2)
Free flow	< 7
Restricted flow	7-10
Dense flow	11-16
Very dense flow	17-26
Jammed flow	> 26

Since we adopt the block-based feature extraction strategy, the number of blocks is an important factor that influences the classification quality. We want to show that the appropriate block number is necessary to obtain better classification results. The results are presented in Table 2. As the observation in our previous work, it can be seen that the classification rate is improved generally as the number of blocks increases. However, a basic principle states that the number of blocks should be large enough so that the texture


(a) Free flow
(b) Restricted flow
(c) Dense flow
(d) Very dense flow

Figure 2. Images with different crowd density levels in PETS 2009

Table 2. Classification rate (%) with different block numbers

Block numbers	2×2	3×3	4×4	5×5
Accuracy	92.50	92.63	93.43	94.98
Block numbers	6×6	7×7	8×8	9×9
Accuracy	93.80	94.10	93.83	92.95

Table 3. Classification rate (%) of different methods

Methods	Classifier	Accuracy
Uniform LBP [15]	SVM	85.7
LBP + PCA + LDA [6]	SVM	89.8
LBP + Gabor [7]	SVM	90.3
TFP [24]	SVM	92.7
Ours	SVM	94.9

features can be represented reliably. On the other hand, it should be small enough to accurately describe the local textures of a crowd image [28]. From the experimental result, in order to satisfy both classification accuracy and computation effort requirements, the appropriate number of blocks is 5×5 which can provide better classification performance.

In order to justify the performance of the proposed approach, we compare our approach with other texture-based methods in [15], [6], [7], and [24]. An LBP is called uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa when the corresponding bit string is considered circular [15]. The uniform LBP can reduce the high dimension of the original LBP feature vector. In [6], the principal component analysis and linear discriminant analysis are used to enhance the discriminative and descriptive power of LBP features. In [7], LBP and Gabor features are computed separately over different crowd images and concatenated for estimating the crowd density. The basic concept behind the Taylor feature pattern (TFP) [24] is that it combines the Taylor expansion and LBP. The experimental results are shown in Table 3. It can be observed that although LBP features are good texture descriptors, they fail to characterize the crowd image sufficiently. If we combine other texture features with the LBP, the discriminative ability of LBP will be enhanced. However, the number of patterns of LBC is considerably smaller than that of LBP when using the same number of neighboring pixels. Consequently, the performance of the LBC-based approach is comparable to LBP-based one in terms of classification accuracy with much less execution time and computational complexity. It can be seen that all the methods perform well on classifying crowd density levels, but our approach provides significant higher accuracy than any other feature.

4. **Conclusions.** Crowd density estimation has drawn a lot of attention in intelligent visual surveillance applications for crowd monitoring and management. Among the existing proposed methods for crowd density estimation, using the texture feature to characterize the crowd density in the image is a simple but effective approach. In this paper, we integrate the Taylor expansion and the block-based local binary count pattern to extract the texture features for crowd density estimation. Evaluations on PETS 2009 dataset demonstrate that the proposed approach can obtain a significant improvement in classification performance. Further work of integrating other local texture patterns into our approach is in progress.

Acknowledgment. This work is partially supported by the Ministry of Science and Technology (MOST), Taiwan, under grant MOST 107-2221-E-390-019-MY2. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] S. A. M. Saleh, S. A. Suandi and H. Ibrahim, Recent survey on crowd density estimation and counting for visual surveillance, *Engineering Applications of Artificial Intelligence*, vol.41, pp.103-114, 2015.

[2] M. S. Saleem, M. J. Khan, K. Khurshid and M. S. Hanif, Crowd density estimation in still images using multiple local features and boosting regression ensemble, *Neural Computing and Applications*, doi:10.1007/s00521-019-04021-2, 2019.

[3] A. N. Marana, S. A. Velastin, L. F. Costa and R. A. Lotufo, Automatic estimation of crowd density using texture, *Safety Science*, vol.28, no.3, pp.165-175, 1998.

[4] X. Wu, G. Liang, K. K. Lee and Y. Xu, Crowd density estimation using texture analysis and learning, *Proc. of 2006 IEEE International Conference on Robotics and Biomimetics*, Kunming, China, pp.214-219, 2006.

[5] W. Ma, L. Huang and C. Liu, Crowd density analysis using co-occurrence texture features, *Proc. of the 5th International Conference on Computer Sciences and Convergence Information Technology*, Seoul, South Korea, pp.170-175, 2010.

[6] H. Fradi, X. Zhao and J.-L. Dugelay, Crowd density analysis using subspace learning on local binary pattern, *Proc. of 2013 IEEE International Conference on Multimedia and Expo Workshops*, San Jose, CA, USA, pp.1-6, 2013.

[7] A. K. Pai, A. K. Karunakar and U. Raghavendra, A novel crowd density estimation technique using local binary pattern and Gabor features, *Proc. of the 14th IEEE International Conference on Advanced Video and Signal Based Surveillance*, Lecce, Italy, pp.1-6, 2017.

[8] K. Xu, X. Chen, S. Jing, Y. Liu and C. Kang, A real time hand gesture recognition method based on the multi-class SVM, *ICIC Express Letters, Part B: Applications*, vol.8, no.5, pp.851-857, 2017.

[9] S. W. Sidehabi, A. Suyuti, I. S. Areni and I. Nurtanio, Grading for ripeness of passion fruit using multi-class support vector machine, *ICIC Express Letters, Part B: Applications*, vol.9, no.10, pp.991-998, 2018.

[10] X. Wang, M. Liu, S. Chen and H. Liu, Identification of continuous rotary motor based on improved particle swarm optimizing support vector machine model, *International Journal of Innovative Computing, Information and Control*, vol.14, no.6, pp.2189-2202, 2018.

[11] C. Dewi and R.-C. Chen, Random forest and support vector machine on features selection for regression analysis, *International Journal of Innovative Computing, Information and Control*, vol.15, no.6, pp.2027-2037, 2019.

[12] X. Li, L. Shen and H. Li, 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.

[13] H. Yang, H. Su, S. Zheng, S. Wei and Y. Fan, The large-scale crowd density estimation based on sparse spatiotemporal local binary pattern, *Proc. of 2011 IEEE International Conference on Multimedia and Expo*, Barcelona, Spain, pp.1-4, 2011.
[14] Z. Ye, J. Wang, Z. Wang and H. Lu, Multiple features fusion for crowd density estimation, *Proc. of 2011 the 4th International Conference on Internet Multimedia Computing and Service*, Wuhan, China, pp.42-45, 2012.
[15] T. Ojala, M. Pietikäinen and T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.24, no.7, pp.971-987, 2002.
[16] M. Pietikäinen, A. Hadid, G. Zhao and T. Ahonen, *Computer Vision Using Local Binary Pattern*, Springer-Verlag London Limited, 2011.
[17] Z. Wenyin and F. Y. Shih, Semi-fragile spatial watermarking based on local binary pattern operators, *Optics Communications*, vol.284, nos.16-17, pp.3904-3912, 2011.
[18] C.-C. Lai and C.-H. Ko, Facial expression recognition based on two-stage features extraction, *Optik*, vol.125, no.22, pp.6678-6680, 2014.
[19] L. Liu, P. Fieguth, G. Zhao, M. Pietikäinen and D. Hu, Extended local binary patterns for face recognition, *Information Sciences*, vols.358-359, pp.56-72, 2016.
[20] C. Singh, E. Walia and K. P. Kaur, Color texture description with novel local binary patterns for effective image retrieval, *Pattern Recognition*, vol.76, pp.50-68, 2018.
[21] Y. Zhao, D.-S. Huang and W. Jia, Completed local binary count for rotation invariant texture classification, *IEEE Trans. Image Processing*, vol.21, no.10, pp.4492-4497, 2012.
[22] B. D. Rigling and R. L. Moses, Taylor expansion of the differential range for monostatic SAR, *IEEE Trans. Aerospace and Electronic Systems*, vol.41, no.1, pp.60-64, 2005.
[23] S. Tang, H. Deng, P. Guo, Y. Li and L. Zhang, Generalized PFA for air-missile borne bistatic forward-looking beam-steering SAR with accelerations, *IEEE Access*, vol.6, pp.74616-74627, 2018.
[24] Y. Ding, Q. Zhao, B. Li and X. Yuan, Facial expression recognition from image sequence based on LBP and Taylor expansion, *IEEE Access*, vol.5, pp.19409-19419, 2017.
[25] O. Chapelle, P. Haffner and V. N. Vapnik, Support vector machines for histogram-based image classification, *IEEE Trans. Neural Networks*, vol.10, no.5, pp.1055-1064, 1999.
[26] *CUDA: GPU-accelerated LIBSVM*, Software available at https://github.com/MKLab-ITI/CUDA.
[27] *PETS 2009 Benchmark Data*, Available at http://cs.binghamton.edu/∼mrldata/pets2009.
[28] C.-C. Lai, C.-H. Wu, S.-T. Pan, S.-J. Lee and B.-H. Lin, Gender recognition using local block difference pattern, in *Advances in Intelligent Information Hiding and Multimedia Signal Processing*, J.-S. Pan, P.-W. Tasi and H.-C. Huang (eds.), Springer International Publishing AG 2017, vol.2, 2017.