Research Journal of Applied Sciences, Engineering and Technology 11(7): 740-745, 2015
DOI: 10.19026/rjaset.11.2033
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: April 10, 2015 Accepted: April 22, 2015 Published: November 05, 2015

Research Article

Sub-difference Image of Curvelet Transform for Crowd Estimation: A Case Study at the Holy Haram in Madinah

[1, 2]Adel A. Hafeez Allah, [1]Syed A. Abu-Bakar and [2]Wasim A. Orfali
[1]Department of Electronics and Computer Engineering, Universiti Teknologi Malaysia, Johor, Malaysia
[2]Department of Electrical Engineering, Taibah University, Madinah, Saudi Arabia

Abstract: Counting people and estimating their densities over a certain area is a fundamental task for many artificial intelligence systems. In this study, sub-difference images of curvelet transform are postulated as an efficient source for effective crowd estimation features. The new algorithm is described in detail in the form of a case study conducted at the Holy Haram in Madinah. The application of the difference images extracted by curvelet transform is thus proven to be efficient and useful for further studies. In addition, the proposed method is independent of any background modeling or background subtraction techniques. The method can also handle crowds of different sizes and strong perspective distortion conditions. The estimation procedure is performed using two versions of difference images generated by forward and customized inverse curvelet transforms. The proposed algorithm is then compared with normal difference image utilization.

Keywords: Crowd estimation, curvelet transform, difference image, Holy Haram

INTRODUCTION

Crowd estimation is a crucial and challenging task in many computer vision applications. The accurate count or estimation of the density of people over a certain area is a key indicator of operational and security management. In crowd management and control, normal operating conditions must be maintained. Overcrowding may be an indicator of congestion, delay, or other security abnormalities and rioting; by contrast, low crowds indicate complications or other discomfort zones. People from all over the world visit Holy Harams for the purpose of worship. Thus, ensuring people's comfort during their prayers is a major management objective. The Holy Haram in Madinah has over 3 million visitors a year. It measures over 98,000 m^2 and has more than 42 multi-door entrances. Therefore, ensuring smooth flow at all zones and entrances is a challenging task. To facilitate the distribution of up to 167,000 worshippers throughout the Holy Haram at a time, an intelligent application must be developed to help provide the required level of safety and comfort. Closed-Circuit Television Systems (CCTV) have a basic setup that captures crowd images and transmit them to monitors. These images are assessed by an observer. However, such a routine is tedious. The attention of observers is likely to wander over time and final judgments are usually subjective. Therefore, automated surveillance and reporting is highly beneficial to crowd monitoring.

Image processing techniques for crowd counting may be grouped into three main categories, namely, detection, movement clustering-based and mapping-based methods (Hashemzadeh *et al*., 2013; Ali *et al*., 2013). The first two approaches detect humans or the independent motions of humans over time and either count or relate these data to the final results. By contrast, the third approach counts the number of people without segmenting the foreground or applying any human detection technique. Thus, this method is feasible for crowd estimation (Loy *et al*., 2013). The third type of approach estimates the number of people in a scene based on features extracted from the foreground. Background subtraction or background modeling is the main component of most mapping-based approaches.

The establishment of estimation using the mapping-based methods was first proposed by Davies *et al*. (1995). In this mapping method, two features were considered in counting, namely, the number of pixels of a foreground area and the total number of perimeter edges. A linear relationship was constructed to estimate the number of people in scenes. Since then, many studies have extracted and evaluated new features, including holistic and local ones. Features with holistic textures generated based on statistics are reported in Rahmalan *et al*. (2006) and Chan *et al*. (2008). These features include Gray-Level Co-Occurrence Matrix (GLCM)-based features (Marana *et al*., 1998) and wavelet transform-based features

Corresponding Author: Syed A. Abu-Bakar, Department of Electronics and Computer Engineering, Universiti Teknologi Malaysia, Johor, Malaysia

Fig. 1: Significant perspective distortion in the images of the dataset of the Bab Assalam gate at the Holy Haram in Madinah. An object that is close to the camera occupies more pixels (large bounding box) and contains more detail than a distant object (smaller bounding box)

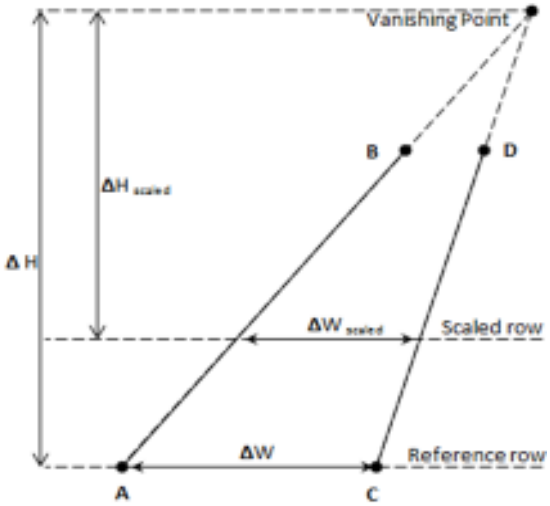


Fig. 2: Calculation of a density map. AB and CD correspond to parallel lines on the ground plane. D is the width of the row and H is the vertical distance from the vanishing point

(Xiaohua *et al*., 2006). Local features include foreground blob size (Kong *et al*., 2006), the histograms of edge orientations (Chan *et al*., 2008), local binary patterns (LBPs) (Ma *et al*., 2008) and feature points such as corners (Albiol *et al*., 2009).

The majority of available map-based techniques requires a reference image to read a scene (Chen, 2013). This requirement is a major drawback of these approaches. An effective model and algorithm must be established to facilitate background and shadow removal, as well as to achieve precise binarization. This binarization need to overcomes the issues of occlusion and of broken blobs, which complicate background subtraction (Zhu *et al*., 2014). A previous study (Chen, 2013) considers frame differences for spatial-temporal features. The method was background-independent, but the resultant difference image is uninformative. Thus, weak features are generated. Another work (Narasimhan *et al*., 2012) utilized a special difference image. However, this image was specific to the study and could not be extended. Other studies attempted to extract a foreground mask to dominate the feature points that relate to the foreground alone (Zhu *et al*., 2014). The work in Hafeez Allah *et al*. (2014) was able

Fig. 3: Normalized density map for the Region of Interest (ROI) in the Holy Haram dataset

to produce an enhanced version of the difference image but did not face the problem of high crowd in high perspective distortion situation. Moreover, there was no illumination problem in the dataset and thus, not to deal with intensity problem.

Perspective correction: As shown in Fig. 1, perspective distortion is a natural phenomenon in which distant objects in an image frame appear smaller than objects that are close to the camera. This distortion was not considered in various previous works (Davies *et al*., 1995; Marana *et al*., 1998; Rahmalan *et al*., 2006) because it is unnoticed, given that the images were taken using high-angle cameras. In cases of severe distortions, however, dependence on any pixel-based feature that relies on the number of foreground pixels or on their intensities will be limited as remote objects occupy fewer pixels than those that are closer to the camera.

Perspective distortion was studied in Ma *et al*. (2004) by computing a density map based on a vanishing point concept. The camera is assumed to be horizontally oriented so that the same weight is assigned to all pixels in the same row of the image matrix, as shown in Fig. 2. A reference row with a unity density is selected in the image; all other rows are scaled with respect to this row. A similar methodology is employed in Chan *et al*. (2008), but object sizes (height multiplied by width) are assigned on the basis of their location in the image. These objects are interpolated over the rest of the image. The resultant matrix represents the weight distortion caused by the perspective problem. A normalization matrix produces the correction density map shown in Fig. 3.

Curvelet transform: Independence from background modeling is one of the aims of the proposed approach on the basis of the assumption that within a short time period, changes to the background are minimal and can be considered unchanged. Hence, the proposed algorithm is influenced by a difference image derived from two sequential frames. However, subtracting two sequential frames removes the background and is unlikely to produce a simplified version of the foreground image.

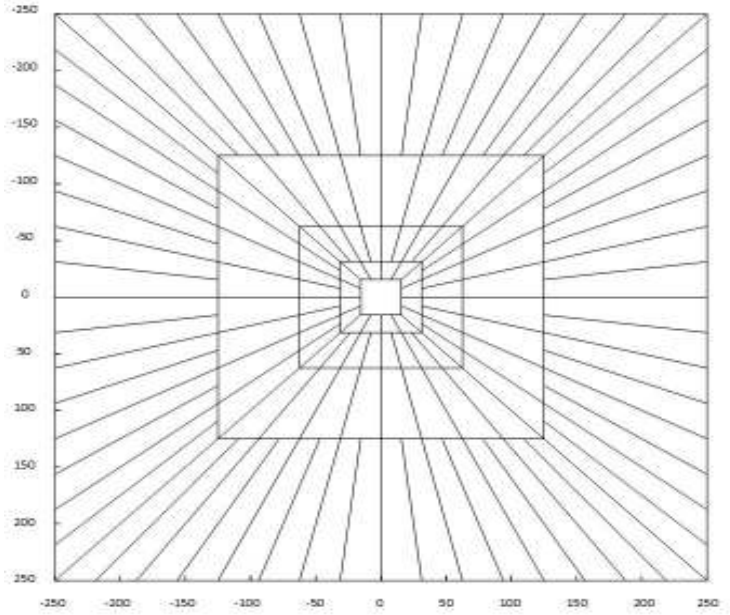


Fig. 4: Curvelet tiling with five scales. The shaded wedge denotes the frequency response of a curvelet at orientation 4 and scale 4

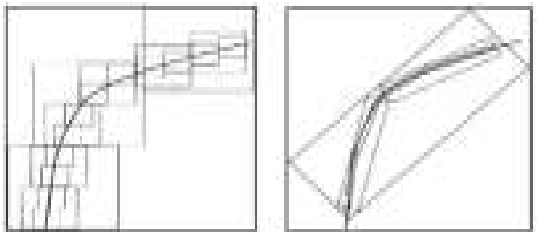


Fig. 5: Representation of edges with many wavelets (left) and few curvelet coefficients (right)

Thus, a transformation process that may preserve more information in the difference image must be developed.

Candes *et al*. (2006) present a new multi-scale transform called curvelet transform. This method is designed to represent curved edges more efficiently than other transforms can, such as wavelet transform. Curvelet transform may decompose an image f of size $N \times N$ for up to $log_2(N) - 3$ scales (levels). These scales contain a different number of orientations. The curvelet frequency tiling with five scales are shown in Fig. 4.

The discrete curvelet transform C^Drepresentation of an image $f$of size $M \times N$ is defined as follows, where $f \in L^2(R^2)$ (Nayak *et al*., 2012):

$$C^D(j,l,k_1,k_2) = \sum_{\substack{0\le m<M \\ 0\le n<N}} f[m,n]\varphi^D_{j,l,k_1,k_2}[m,n] \quad (1)$$

where,

j = The angle
l = The direction
k_1 and k_2 = The spatial locations of the output

Multi-resolution analysis is a basic feature of both wavelet and curvelet transforms. Unlike the wavelet that displays decompositions at every π/4 angle, angles are doubled at different scales in curvelet transform (Majumdar and Nasiopoulos, 2008). Thus, this method induces a fine directional decomposition. The curvelet can represent curves using a few coefficients, as shown in Fig. 5. The additional angles enhance curve representation, especially given curve-like edges (Jianwei and Plonka, 2010). These edges and represented corners are considered in Kausalya and Chitrakala (2012) and Bahashwan and Abu Bakar (2015). Thus, the hypothesis of this study is that using curvelet transforms can aggravate the discontinuity problem of curved edges (Guha and Wu, 2010). Moreover, the additional directions increase the containment of the curvelet coefficients of pixels belonging to a particular object.

APPROACH

The current paper proposes a new method for crowd estimation based on a customized difference image obtained by curvelet transform. The new method applies the transform to two sequential frames and employs different customized inverse curvelet transforms with distinct difference scales, as in the system architecture depicted in Fig. 6.

Pre-processing: The case study is conducted it Bab Assalam gate which naturally close to a bright outdoor area. Thus, the resultant high intensity values for gate-close objects can affect the final difference images and consequently, the extracted features. Hence, a

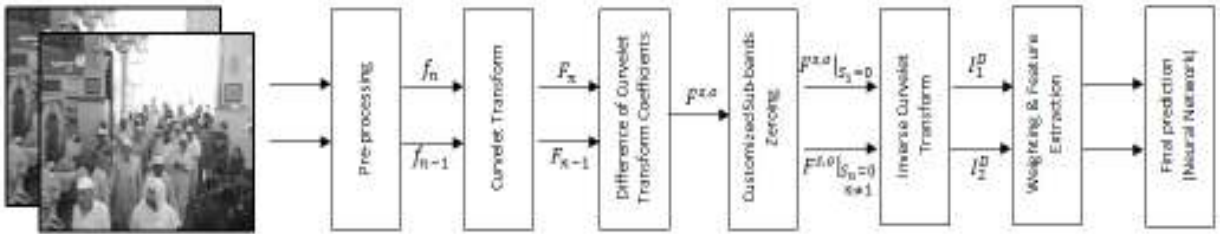
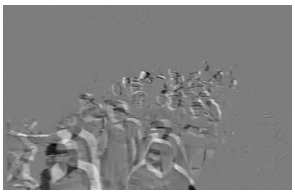


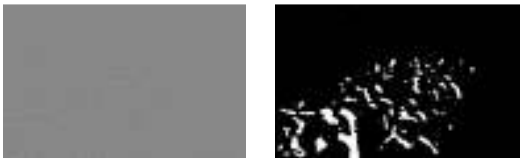
Fig. 6: System architecture

(a) (b)



(c)

Fig. 7: (a) Original grayscale frames, (b) normal difference image and (c) difference image obtained using the curvelet transform



(a) (b)

Fig. 8: Customized inverse curvelet transform; (a) I_1^D and (b) binarized I_2^D

preprocessing step must be conducted with a normalizing algorithm. Our approach employs a histogram equalization to improve contrast. This equalization non-linearly maps the intensity values to expand the range of intensities in the image.

Difference image with customized curvelets: The hypothesis of this study is that using the difference image derived from two sequential frames via curvelet transform enhances the difference image. Figure 7 illustrates a comparison between the normal difference images extracted from two grayscale frames from our case study dataset and those obtained by the use of the curvelet transform. This difference image is derived by applying the curvelet transform to two sequential frames and by subjecting the decomposed difference to inverse curvelet transform. The difference image between the two frames produces a silhouette of the moving objects. This silhouette image is useful for further calculations involving moving objects alone. Hence, any improvements to such difference images enhances results. This outcome is the main advantage of using curvelet transform. However, different scales in the curvelet-decomposed images preserve various types of information. Applying customized inverse curvelet transform to the chosen scales can enhance feature extraction. Moreover, the new sub-scale difference images generate additional information through the supply of accurate features given that curvelet transform preserves more accurate information.

The proposed algorithm is shown in Fig. 7. The algorithm first decomposes two consecutive frames ($f_{n-1}$and f_n) using curvelet transform to produce F_{n-1} and F_n, respectively. These variables are the representations of the decomposed curvelet domain image obtained from the two frames. The number of decomposed images depends on the number of scales and orientations in curvelet transform. Second, the difference between the two curvelet coefficients is determined. A simplified version of the foreground in its new band $F^{s,o}(i,j)$ may be generated by subtracting the common regions in all decomposed curvelet transform coefficients:

$$F^{s,o} = F_n^{s,o} - F_{n-1}^{s,o} \quad (2)$$

where,

S = The scale of the curvelet decomposed image
O = The orientation within this scale

Rather than obtaining a difference image directly, the algorithm performs two customized inverse transforms. The first transformation discards the coefficients for first scale ($S = 1$) alone, whereas the second discards the other scales ($S \neq 1$). The two resultant difference images DI_1 and $DI_2$are employed as sources for feature extraction.

Feature extraction: The following features are extracted from relevant customized inverse curvelet transform images. Figure 8 shows a sample $I_1^D$and a binarized I_2^D. To overcome the perspective distortion effect, the density map is used to weight the customized

Table 1: Level of service

Level of service	Range of density (people per unit area)	Number of people	Group
Restricted flow	0.5-0.80	9-16	Low
Dense flow	0.81-1.26	17-25	Moderate
Very dense flow	1.27-2.0	26-32	High
Jammed	>2.0	>33	Very high

Table 2: Final classification for both systems

		Curvelet difference images		Normal difference images	
Range of people (class)	No. of frames in class	True detection	True detection (%)	True detection	True detection (%)
Low	192	192	100	26	13.5
Moderate	210	147	70.0	110	50.0
High	429	394	91.8	265	61.8
Very high	269	268	99.6	97	32.4

difference images generated at a certain level. The first difference image I_1^D is utilized as a source for extracting edges and corners. The total number of edges and corners detected in the edge image is weighted and then featured. By contrast, the weighted binarized version of the second difference image I_2^D produces a relative area feature. The final classification is the result of using a back-propagation neural network. Three equivalent features are applied to the comparison test, namely, the binarizationblob sizes, total edges and corner points. However, this test is conducted on the normal difference images alone.

Density calculation and classification: Many methods have been developed to estimate crowd density. The estimation may be represented in the form of an exact count of a number of people per unit square, in percentage levels, or even in density classes. The level of service defined by Polus *et al*. (1983) classifies crowds according to the occupied area. On the basis of the range of average area occupancy, five groups of densities (namely, very low, low, moderate, high and very high crowd) are defined. However, our data set does not include very low crowd images. Thus, this group density has been ignored as shown in Table 1.

EXPERIMENTAL RESULTS

The proposed algorithm incorporates the enhanced difference images using curvelet transform. To illustrate the enhanced and the final results, two main comparison tests are conducted for evaluation, namely, crowd estimation based on the proposed algorithm and a second test based on normal difference images. For the proposed algorithm, four scales are used for curvelet transform. These scales consist of 1, 8 and 16 angles for the first, second and third scales, respectively and of one fine level at the final scale. Both tests are performed with the Holy Haram dataset, which contains 1,100 frames with five frames per second. Each frame is 720×576 pixels. Table 2 illustrates the true classification of the system in comparison with actual ground truth for all tests. Figure 9 compares the true and false classifications among classes.

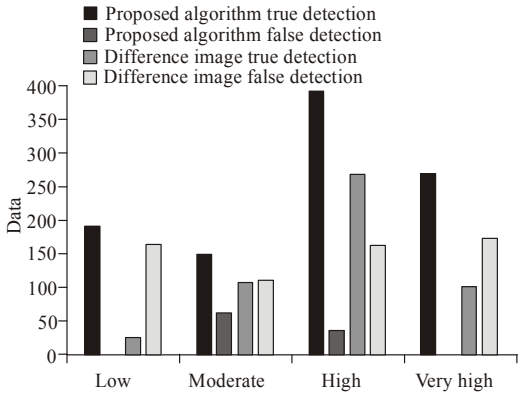


Fig. 9: Chart that compares true, false and missed detection rates

The overall system accuracy of the proposed system is 91% when curvelet transform is utilized. The accuracy obtained with the normal difference image is 45%. The calculation time required by the testing system for extracting features from a sub-difference image depends on frame size. The time required by both forward and inverse curvelet transforms is 1.125 sec, whereas the system for normal difference images requires 0.37 sec.

CONCLUSION

This study presents a method that is independent of background modeling. This new method can handle the perspective distortion problem and its main advantage is its capability to extract an enhanced version of the difference image. Curvelets can customize difference images into different versions with varied features for further calculation. Thus, the proposed method is suitable for situations wherein background modeling techniques are lacking and wherein scene backgrounds change constantly. Curvelet difference images are informative and generate good features that are not limited to those presented in this study. This success encourages the use of the different versions of decomposed curvelet images.

REFERENCES

Albiol, A., M.J. Silla, A. Albiol and J.M. Mossi, 2009. Video analysis using corner motion statistics. Proceeding of the IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, pp: 31-38.

Ali, S., K. Nishino, D. Manocha and M. Shah, 2013. Modeling, Simulation and Visual Analysis of Crowds: A Multidisciplinary Perspective. Springer, NY, ISBN: 978-1461484820.

Bahashwan, M.A. and S.A.R. Abu Bakar, 2015. Offline handwritten Arabic character recognition using features extracted from curvelet and spatial domains. Res. J. Appl. Sci. Eng. Technol., 11(2): 158-164.

Candes, E., L. Demanet, D. Donoho and L. Ying, 2006. Fast discrete curvelet transforms. Multiscale Model. Sim., 5(3): 861-899.

Chan, A.B., Z.S. Liang and N. Vasconcelos, 2008. Privacy preserving crowd monitoring: Counting people without people models or tracking. Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR, 2008), pp: 1-7.

Chen, J., 2013. Crowd counting based on difference images. Electron. Electr. Eng., 19(2): 83-87.

Davies, A.C., J.H. Yin and S.A. Velastin, 1995. Crowd monitoring using image processing. Electron. Commun. Eng., 7(1): 37-47.

Guha, T. and Q.M.J. Wu, 2010. Curvelet based feature extraction. Face Recogn., pp: 35-46.

Hafeez Allah, A.A., S.A. Abu Bakar and W.A. Orfali, 2014. Curvelet transform sub-difference image for crowd estimation. Proceeding of the IEEE International Conference on Control System, Computing and Engineering (ICCSCE, 2014), pp: 502-506.

Hashemzadeh, M., G. Pan and M. Yao, 2013. Counting moving people in crowds using motion statistics of feature-points. Multimed. Tools Appl., 72(1): 453-487.

Jianwei, M. and G. Plonka, 2010. The curvelet transform. IEEE Signal Proc. Mag., 27(2): 118-133.

Kausalya, K. and S. Chitrakala, 2012. Idle object detection in video for banking ATM applications. Res. J. Appl. Sci. Eng. Technol., 4(24): 5350-5356.

Kong, D., D. Gray and H. Tao, 2006. A viewpoint invariant approach for crowd counting. Proceeding of the 18th International Conference on Pattern Recognition (ICPR, 2006), pp: 1187-1190.

Loy, C.C., K. Chen, S. Gong and T. Xiang, 2013. Crowd counting and profiling: Methodology and evaluation. In: Ali, S., K. Nishino, D. Manocha and M. Shah (Eds.), Modeling, Simulation and Visual Analysis of Crowds. Springer Science+Business Media, New York, 11: 347-382.

Ma, R., L. Li, W. Huang and Q. Tian, 2004. On pixel count based crowd density estimation for visual surveillance. Proceeding of the IEEE Conference on Cybernetics and Intelligent Systems, pp: 170-173.

Ma, W., L. Huang and C. Liu, 2008. Advanced local binary pattern descriptors for crowd estimation. Proceeding of the Pacific-Asia Workshop on Computational Intelligence and Industrial Application (PACIIA'08), pp: 958-962.

Majumdar, A. and P. Nasiopoulos, 2008. Frontal face recognition from video. In: Bebis, G. *et al*. (Eds.), ISVC 2008, Part II, LNCS 5359, Springer, Berlin, Heidelberg, pp: 297-306.

Marana, A., S. Velastin, L.D.F. Costa and R. Lotufo, 1998. Automatic estimation of crowd density using texture. Safety Sci., 28(3): 165-175.

Narasimhan, K., V. Elamaran, S. Kumar, K. Sharma and P.R. Abhishek, 2012. Comparison of satellite image enhancement techniques in wavelet domain. Res. J. Appl. Sci. Eng. Technol., 4(24): 5492-5496.

Nayak, R., J. Bhavsar, J. Chaudhari and S.K. Mitra, 2012. Object tracking in curvelet domain with dominant curvelet subbands. Energy, 16(775.86): 32.

Polus, A., J.L. Schofer and A. Ushpiz, 1983. Pedestrian flow and level of service. J. Transp. Eng-ASCE, 109(1): 46-56.

Rahmalan, H., M.S. Nixon and J.N. Carter, 2006. On crowd density estimation for surveillance. Proceeding of the Institution of Engineering and Technology Conference on Crime and Security, pp: 540-545.

Xiaohua, L., S. Lansun and L. Huanqin, 2006. Estimation of crowd density based on wavelet and support vector machine. T. I. Meas. Control, 28(3): 299-308.

Zhu, Y., R. Liang and H. Wang, 2014. Counting crowd flow based on feature points. Neurocomputing, 133: 377-384.