Using Facial Expression Recognition for Crowd Monitoring

Ross Philip Holder and Jules-Raymond Tapamo[✉]

School of Engineering, University of KwaZulu-Natal, Durban 4041, South Africa
tapamoj@ukzn.ac.za

Abstract. In recent years, Crowd Monitoring techniques have attracted emerging interest in the field of computer vision due to their ability to monitor groups of people in crowded areas, where conventional image processing methods would not suffice. Existing Crowd Monitoring techniques focus heavily on analyzing a crowd as a single entity, usually in terms of their density and movement pattern. While these techniques are well suited for the task of identifying dangerous and emergency situations, they are very limited when it comes to identifying emotion within a crowd. In this work, we propose a novel Crowd Monitoring algorithm based on estimating crowd emotion using Facial Expression Recognition (FER). By isolating different types of emotion within a crowd, we aim to predict the mood of a crowd even in scenes of non-panic. To validate the effectiveness of the proposed algorithm, a series of cross-validation tests are performed using a novel Crowd Emotion dataset with known ground-truth emotions. The results show that the algorithm presented is able to accurately and efficiently predict multiple classes of crowd emotion even in non-panic situations where movement and density information may be incomplete.

1 Introduction

Crowd Monitoring is a topic of emerging interest in the field of computer vision and was born largely from the desire to monitor the nature of groups of individuals in crowded areas, where conventional image processing methods would not suffice [31]. Areas where Crowd Monitoring systems are commonly deployed include airport terminals, sports stadiums, and other public facilities that attract large crowds of people. Crowd Monitoring can be used to aid law enforcement in recognizing and identifying crowds that may cause public disorder. Examples include identifying disorderly crowds of sports fans that may have gathered after a football match, or a group of disgruntled protesters that have taken to the street. With the advent of social media platforms, such as Twitter, small gatherings can often gather momentum very quickly, evolving into large crowds that can be difficult to control [5]. This necessitates the need for advances in Crowd Monitoring techniques.

Facial Expression Recognition (FER) [18,21,23] is a technique used to extract and classify emotion from an individual's facial expression. It is widely accepted

that there are seven universally recognizable emotions as first identified by Ekman [12], namely: joy, surprise, anger, fear, disgust, sadness and neutral emotion. In this work we use FER to extract and classify emotion from individuals in a crowded environment. The individual emotions can be combined to estimate the emotion of the crowd.

Due to the difficulty associated with extracting individuals from a crowd, most Crowd Monitoring techniques focus heavily on analyzing crowds as a single entity. Many different holistic based [2,3,6,10,30] and object-level based [7,8,24, 32] methods of Crowd Monitoring have been proposed in current literature, such as analyzing crowd movement patterns, flow and density. While these approaches are well suited for the task of identifying emergency situations, such as a large group of people exiting a building at once or a crowd gathering around a fight, they are very limited when it comes to identifying the nature or mood of a crowd outside of scenes of panic. A system that is able to autonomously identify the mood of a crowd in real-time dynamic environments is required.

There is potential for aggressive crowds, fueled by their sense of superiority in numbers [9], to vandalize and loot property while endangering the lives of innocent bystanders. By identifying the mood of a crowd in real-time, the system can help to alert officials to potentially aggressive and disorderly crowds so that necessary measures, such as additional policing units, can be deployed to prevent further aggression and violence. In areas where policing units are limited, the system allows officials to concentrate available units on crowds of interest; maximizing their resources and efficiency. The system uses emotion to represent the mood of the crowd. Crowd emotion can be estimated at object-level using FER.

2   Materials and Methods

This section presents methodology for estimating the overall emotion of a crowd. Firstly, the popular Viola and Jones face detection algorithm is used to detect and extract unobscured faces from individuals in the crowd. Next, a robust and efficient method of FER is used together with a machine learning algorithm to extract and classify each facial expression as one of seven universally accepted emotions [12]. Finally, the emotion of the crowd is estimated by isolating groups of similar emotion based on their relative size and weighting.

2.1   Face Detection

The Viola and Jones [28] face detection algorithm, which uses a boosted cascade of classifiers to rapidly detect faces, has been shown to be extremely effective at identifying faces in uncontrolled backgrounds with great accuracy [17] compared to other existing face detection techniques. In our work, the Viola and Jones method was selected for face detection due to its combination of speed and accuracy. The Viola and Jones face detection algorithm consists of three main steps: (1) Computing the integral image, (2) Learning classifiers using Adaboost, and (3) Combining the classifiers in a cascade structure.

2.1.1   Computing the Integral Image

Images are classified using simple features as opposed to pixel intensities. The simple features used are reminiscent of Haar Basis functions and consist of two, three and four rectangle features. Because the set of rectangle features can be very large, the images are first represented by an integral image. The integral image at location (x, y) represents the sum of the pixels above and to the left of (x, y), inclusive:

$$ii(x, y) = \sum_{x\prime \leq x, y\prime \leq y} i(x\prime, y\prime) \tag{1}$$

where $ii(x, y)$ is the integral image and $i(x, y)$ is the original image. By using the integral image, the time taken to compute the rectangular feature set at any scale or location is greatly reduced because any rectangular sum can be computed using just four array references.

2.1.2   Learning Classifiers Using Adaboost

The number of rectangle features associated with each image sub-window is far greater than the number of pixels. To ensure fast classification, only a small subset of these features are combined to form an effective classifier. Adaboost [13] is used in such a way that each weak learning algorithm selects only a single rectangle feature which best separates the positive and negative examples. For each of these features, the optimal threshold classification function is computed such that the minimum number of examples are misclassified. A weak classifier $h_j(x)$ is thus represented by:

$$h_j(x) = \begin{cases} 1, \text{ if } p_j f_j(x) < p_j \theta_j \\ 0, \text{ otherwise} \end{cases} \tag{2}$$

where f_j is a feature, θ_j is the threshold, p_j is a parity indicating the direction of the inequality and x is a 24×24 pixel sub-window of an image.

2.1.3   Combining the Classifiers in a Cascade Structure

To speed-up the classification process, successively more complex classifiers are combined in a cascade structure. Each stage in the cascade is constructed by training a classifier using Adaboost with the threshold adjusted to minimize false negatives. By using a cascade of classifiers, sub-windows that are not of interest can be quickly discarded in the early stages so that increased computation is spent only on more promising face-like regions in the later stages; greatly increasing the overall computational efficiency of classification.

2.2   Facial Expression Recognition (FER)

FER consists mainly of three important steps [21]: (1) Pre-processing of facial images, (2) Facial feature extraction, and (3) Expression classification. Due to the wide variety of individuals that can be found in a crowd; an accurate, efficient

and robust method of FER is required for the purposes of Crowd Monitoring. In this work, the detected faces are pre-processed to remove non-discriminative expression regions of the face and Gradient Local Ternary Pattern (GLTP) [1] is applied for facial feature extraction. A Support Vector Machine (SVM) [16] is used for feature classification. Each detected facial expression in the crowd is classified as one of seven universally accepted emotions [12].

2.3   Computing the Distance Between Faces

Before we can find groups of individuals situated close together in the crowd, we first need to determine the distance between neighbouring faces. Each face is treated as a node, where the vertex of the node is represented by the top left point of the region of interest (ROI) representing the face. As in [11], a fully-connected undirected graph is used to link every node's vertex with one another, where the distance between any two nodes is represented by the weight of the connecting edge. We say the resulting graph is fully-connected because each node is connected to every other node present, and undirected because there is only one unique edge between each pair of nodes (direction does not matter). As such, for N nodes we have a total of $(N \times (N-1))/2$ edges; where the distance between nodes i and j is found using the Euclidean norm as:

$$\text{Distance}_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{3}$$

where (x_i, y_i) represents the vertex of node i and (x_j, y_j) represents the vertex of node j. The graph can be represented by an $N \times N$ adjacency matrix (Adj_Mat), where $\text{Adj\_Mat}_{i,j} = \text{Distance}_{i,j}$. The weight of each edge is the Euclidean distance between the nodes. The fully-connected undirected graph for a crowd of 20 people is shown in Fig. 1.
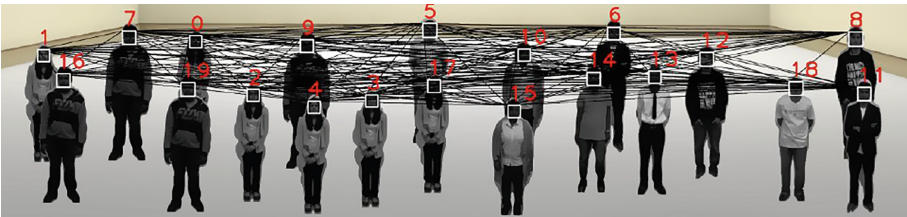

Fig. 1. Fully-connected undirected graph for a crowd of 20 people

2.4   Computing the Closest Neighbours of Each Face

A Minimum Spanning Tree (MST) is used to represent each face's closest neighbours as suggested in [11]. A spanning tree of a graph G is a tree, where every edge in the tree belongs to G and, that includes every node of G. The cost of a

spanning tree is represented by the sum of the weights of all edges in the tree. A MST is a spanning tree where the cost is a minimum. Numerous approaches have been suggested for finding a MST. The two most popular approaches are Kruskal's algorithm and Prim's algorithm [27]. In this work, Prim's algorithm was used to find the MST. Starting with an empty MST, for each step of Prim's algorithm, we consider a group of edges that connects the set of nodes already included in the MST with the set of nodes not yet included. The edge with minimum weight is selected and the node is added to the MST. The procedure is repeated until all nodes have been included in the MST. The MST for the fully-connected undirected graph of the crowd given in Fig. 1 is shown in Fig. 2. In a MST there is a total of $N-1$ edges.
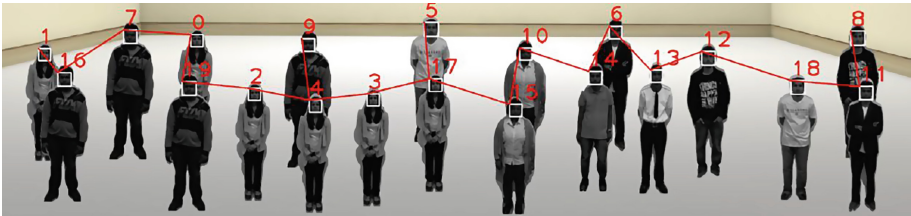

Fig. 2. Minimum spanning tree for a crowd of 20 people

2.5 Estimating Crowd Emotion from Groups of Similar Emotion

The predicted emotion of each face and the MST can be used to identify groups of individuals who are expressing similar emotion and who are situated close together in the crowd. These groups of individuals can be represented by chains of emotion, where the length of each chain is represented by the number of individuals in the chain. The overall emotion of the crowd can then be estimated by finding the largest chain of emotion with the greatest weighting. This approach is more accurate at estimating crowd emotion compared to more simplistic methods such as finding the predominant individual emotion in the crowd. The size of each emotion chain in relation to the crowd is compared to a set threshold value, $thresh$, which represents the minimum size required for the chain to be considered large enough to influence the overall crowd emotion. Each prototypic emotion is assigned a weighting representing its importance. In our work, all emotions are assigned an equal weighting with the exception of neutral emotion which is assigned a lower weighting. This is because neutral emotion does not provide much information about the emotional state of the individuals within the crowd. The overall crowd emotion is predicted as the emotion belonging to the chain that meets the following requirements:

1. The size of the chain in relation to the crowd is greater than or equal to a threshold, $thresh$.

2. The emotion of the chain has the greatest possible weighting out of the chains that meet requirement (1).
3. The size of the chain is the largest out of the chains that meet requirements (1) and (2).

If no chain meets the above requirements; the emotion of the crowd is considered to be mixed. Because individuals in a crowd can take on the emotion of the people around them, it is possible that even a relatively small group of individuals expressing one emotion can influence the emotion of the individuals around them who in turn can influence the individuals around them. This chain reaction is known as the Domino effect and can potentially lead to crowds getting out of control. Our proposed crowd emotion estimation technique aims to identify sufficiently large groups of individuals expressing similar emotion in the crowd, such as anger, before it is able to spread any further. This allows for early detection of potentially problematic crowds.

Consider the crowd given in Fig. 2. The emotion chains for the crowd are illustrated in Fig. 3, where the values above each node represent the node number and predicted FER emotion label of the node. There are a total of 2 unique emotion chains in the crowd; one with emotion label 0 (anger) and another with emotion label 4 (neutral). In this work, the required threshold is set to $thresh = 30\%$ (this value is considered optimal since negative groups of emotion in the crowd can be detected early while false detections are kept to a minimum). The size of both chains are greater than the required threshold. The anger chain has a greater weighting than the neutral chain and because there are no other emotion chains with an equivalent or greater weighting, the overall emotion of the crowd is predicted to be anger.
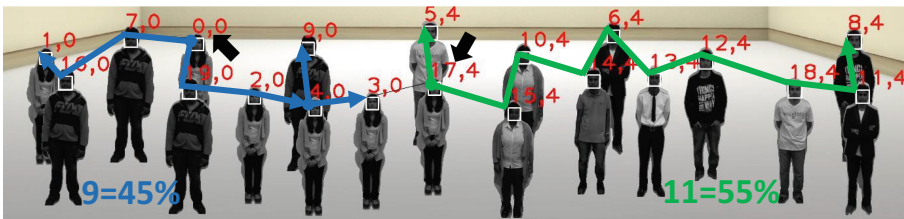

Fig. 3. Finding chains of emotion in the crowd

3   Experimental Setup

In this section, the dataset and procedure used for testing our proposed algorithm are presented.

3.1   Crowd Emotion Dataset

Existing Crowd Monitoring datasets [14, 20, 22, 26, 29] are unsuitable for extracting facial expressions and do not provide known ground-truth emotion labels.

We thus propose the creation of a novel Crowd Emotion dataset with known ground-truth emotion labels. Images from the Extended Cohan-Kanade (CK+) [19] facial expression dataset are pre-processed and placed together in an empty environment to simulate crowd images. The images represent a crowd under optimal conditions with no facial obscurities present. Each crowd image consists of 2 groups of 10 subjects. To produce a ground-truth emotion, subjects in one group are placed so that they are expressing random emotions, none of which exceed the threshold value, while the subjects in the remaining group are placed so that they are expressing the ground-truth emotion. A generated crowd image with ground-truth emotion anger is shown in Fig. 4.
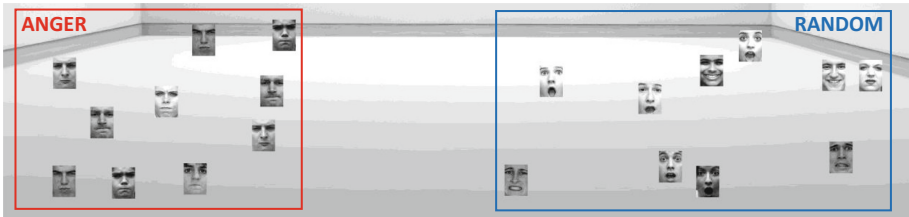

Fig. 4. Generated crowd image with ground-truth emotion anger

3.2   Testing Procedure

To find the average recognition accuracy of our proposed algorithm, we implement a 10-fold cross-validation testing procedure using pre-processed facial images from the CK+ dataset. The images are randomized and divided into 10 roughly equally-sized segments. For each fold, 9 of the segments are used for training the classifier while the remaining segment is used to generate crowd images for testing. This ensures that none of the subjects used for training the classifier are included in the crowd image under test. This process is repeated for the remaining 9 folds and the average recognition accuracy is calculated across all 10 folds.

We define 8 (joy, surprise, anger, fear, disgust, sadness, neutral, mixed), 7 (excludes neutral), and 2 (emotions are grouped into positive and negative) classes of crowd emotion for testing. For 8 & 7 classes of crowd emotion, 3 crowd images are generated for each class per fold, resulting in a total of 240 crowd images for 8 classes and 210 crowd images for 7 classes. For 2 classes of crowd emotion, 12 positive emotion and 12 negative emotion crowd images are generated per fold, resulting in a total of 240 crowd images tested.

4   Results and Discussion

In this section, results are reported on the proposed Crowd Emotion dataset for the algorithm presented.

4.1   Recognition Accuracy

The recognition accuracies achieved for 8, 7, and 2 classes of crowd emotion are summarized in Table 1. An average recognition accuracy of 64.6% was achieved for 8 classes of crowd emotion. Examining the crowd emotion confusion matrix shown in Table 2, we find that joy, neutral and mixed crowd emotions exhibited a high degree of recognition accuracy. On the contrary, anger and sadness emotions exhibited a very poor degree of recognition accuracy. These findings share a direct correlation with the chosen method of FER, which achieved an average recognition accuracy of 85.4% on the crowd images. The confusion matrix for FER is given in Table 3 and shows that out of the 7 facial emotions on test, anger and sadness emotions achieved the lowest recognition accuracies; being confused to a great extent with neutral emotion.

Table 1. Recognition accuracy (%) for 8, 7 and 2 classes of crowd emotion

Classes of crowd emotion	Recognition accuracy (%)
8 class	64.6 ± 2.1
7 class	81.3 ± 1.6
2 class (with neutral emotion)	72.4 ± 2.5
2 class (without neutral emotion)	94.8 ± 1.1

Table 2. Crowd confusion matrix (%) for 8 classes of crowd emotion

	Joy	Sur	Ang	Fear	Dis	Sad	Neu	Mixed
Joy	**97**	0	0	0	0	0	3	0
Sur	0	**62**	0	0	0	0	37	1
Ang	0	0	**5.7**	0	0	0	91.6	2.7
Fear	0	0	0	**58.3**	0	0	38	3.7
Dis	0	0	0	0	**66.3**	0	32	1.7
Sad	0	0	0	0	0	**29**	67	4
Neu	0	0	0	0	0	0	**100**	0
Mixed	0	0	0	0	0	0	1.7	**98.3**

An average recognition accuracy of 81.3% was achieved for 7 classes of crowd emotion. This shows a 16.7% improvement compared to when neutral emotion was included. Examining the crowd emotion confusion matrix in Table 4, we note that while all emotion classes displayed an improvement in recognition accuracy compared to 8 class testing, in particular, anger and sadness emotions experienced the largest improvement; having increased more than threefold. This is

Table 3. FER confusion matrix (%) for 8 classes of crowd emotion

	Joy	Sur	Ang	Fear	Dis	Sad	Neu
Joy	**98.8**	0	0.1	0	0	0	1.1
Sur	1.3	**92.3**	0.3	0.2	1.6	0.1	4.2
Ang	0	3.6	**51.3**	0.7	3.1	1.6	39.7
Fear	1.3	3.2	0.3	**85.1**	0	3.9	6.2
Dis	3	0.5	0.1	0.1	**87.9**	0	8.4
Sad	0	1.2	0.1	0.3	0	**71.5**	26.9
Neu	0.2	1	0.3	0.7	0.5	1.1	**96.2**

Average: 85.4%

supported by the FER confusion matrix given in Table 5, where anger and sadness emotions experienced the most significant increase in recognition accuracy out of the 6 facial emotions on test. With neutral emotion excluded, the average FER recognition accuracy improved by 7.6% from 85.4% to 93%. Further examination of both 7 class and 8 class FER confusion matrices shows that pleasing emotions such as joy and surprise tend to exhibit higher recognition accuracies compared to other displeasing emotions such as anger, fear and disgust, which often get confused between one another. This is evident in Table 5, where anger and fear is confused with disgust and sadness.

Table 4. Crowd confusion matrix (%) for 7 classes of crowd emotion

	Joy	Sur	Ang	Fear	Dis	Sad	Mixed
Joy	**98.3**	0	0	0	0	0	1.7
Sur	0	**79**	0	0	0.3	0	20.7
Ang	0	0	**51.7**	0	0	0	48.3
Fear	0	0	0	**64.3**	0	0	35.7
Dis	0	0	0	0	**85.3**	0	14.7
Sad	0	0	0	0	0	**91.3**	8.7
Mixed	0.3	0	0	0	0.3	0	**99.4**

We reduce the 8 and 7 classes of crowd emotion into just 2 classes - positive and negative. Emotions that can be considered pleasing are grouped into the positive class while emotions that can be considered displeasing are grouped into the negative class. For what was previously 7 classes of crowd emotion, we group joy and surprise into the positive class while anger, fear, disgust and sadness are grouped into the negative class. For what was previously 8 classes of crowd emotion, we consider neutral emotion to be non-negative and place it in

Table 5. FER confusion matrix (%) for 7 classes of crowd emotion

	Joy	Sur	Ang	Fear	Dis	Sad
Joy	**99.5**	0	0.4	0	0.1	0
Sur	1.1	**95.5**	0.6	0.6	2.1	0.1
Ang	0	7	**83**	1	4.4	4.6
Fear	1.7	3	0.9	**86.2**	0.1	8.1
Dis	3.2	0.8	1.7	0	**94.2**	0.1
Sad	0.1	2.3	1.3	0.2	0.1	**96**
Average: 93.0%						

the positive emotion class. Crowd's of mixed emotion are also considered non-negative and thus classified as positive. We repeat our cross-validation testing on the reduced class set for 2 given scenarios: (1) neutral emotion is included as part of the positive emotion class and (2) neutral emotion is excluded.

An average recognition accuracy of (1) 72.4% (neutral emotion included) and (2) 94.8% (neutral emotion excluded) was achieved for 2 classes of crowd emotion. These results show an improvement in accuracy of 7.8% compared to 8-class testing and 13.5% compared to 7-class testing. We note that by excluding neutral emotion from 2 class testing, recognition accuracy improved by 22.4% compared to when it was included. This significant increase in recognition accuracy due to the exclusion of neutral emotion is consistent with our findings during 7 class testing, where we also noted a significant increase in accuracy compared to 8 class testing. The crowd emotion confusion matrices for 2 classes of crowd emotion are given in Tables 6 and 7. In both cases, all crowd images with positive emotion were correctly predicted; demonstrating that positive emotions may be more easily recognized compared to negative emotions. For the first case, with neutral emotion included, more than half of the negative emotion crowd images on test were misclassified. Some negative emotions, such as anger and sadness, would have been misclassified as neutral emotion causing those crowd images to be incorrectly classified as having positive emotion. For the second case, with neutral emotion excluded, the number of crowd images with negative emotion that were correctly predicted was much higher; resulting in the largest average recognition accuracy achieved on test. Overall, these findings show that greater accuracies can be achieved by combining multiple emotions of a similar type to form a reduced class set, while maintaining the ability to discern negative crowd emotion from positive crowd emotion.

**Table 6.** Crowd confusion matrix (%) for 2 classes of crowd emotion (with neutral)

	Positive	Negative
Positive	**100**	0
Negative	55.2	**44.8**

Table 7. Crowd confusion matrix (%) for 2 classes of crowd emotion (without neutral)

	Positive	Negative
Positive	**100**	0
Negative	10.2	**89.8**

4.2 Efficiency

To test the performance of our proposed algorithm, we vary the size of crowd while measuring the average time taken to predict the emotion of each crowd image on a Core 2 Duo, with a clock-speed of 2.0 GHz and 3 GB of RAM. The individuals placed in the crowd are selected at random and the results are given in Fig. 5. The results show a linear relationship between crowd size and prediction runtime. We note that for small crowds of 1 to 20 people, prediction takes less than 1 s. On the other hand, for larger crowds of 200 to 220 people, it takes in the region of 12 to 13 s for each prediction. Overall the algorithm shows potential for real-time application.
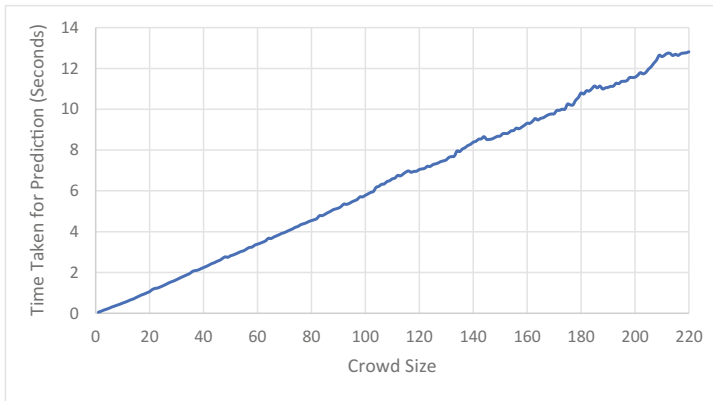
Fig. 5. The effect of varying crowd size on prediction runtime

4.3 Comparison to Results in Literature

We compare our proposed algorithm to existing Crowd Monitoring techniques aimed at emotion detection in crowds. Although a direct comparison cannot be made due to differences in the datasets and the testing procedures used, we outline any advantages and disadvantages between methods and where possible compare accuracies. In [25], it was proposed that emotion-based classification of a crowd could be used to better predict crowd behaviour. The authors created a novel crowd behaviour dataset consisting of video sequences for 5 types of crowd behaviour annotated with 6 emotion labels (disgust was excluded) based on the motion of the crowd. Using dense trajectory and SVM classification, emotion

descriptions were extracted for each video sequence and mapped to a crowd behaviour. The authors reported a recognition accuracy of 43.9% using a leave-one-out testing procedure (which typically gives higher accuracies), 20.7% lower than our 8 class results, although the dataset used in their work was considerably more difficult. Although the authors work represents a novel approach to Crowd Monitoring through the use of crowd emotion, it requires obtaining video sequences of crowds around the apex of their behaviour to be truly effective, which is a complex real-world task. The method is also highly dependent on the type of crowd sequences supplied during the training stage and thus may not work in all environments. In comparison, our proposed method focuses only on 2D static images, which is far more computationally efficient for practical real-world applications. By relying solely on facial expressions for emotion classification, our method should not be greatly effected by changing environments or scenery within the crowd (apart from illumination variation and noise).

In [4], a dynamic probabilistic clustering technique was proposed to model a crowd's response to different events. A simulation model to produce evacuation and panic situations was implemented to test the proposed method. Crowd emotion was classified as either positive or negative based on the clustering together (herding) of individuals within the crowd in response to panic situations. The authors report that a recognition accuracy of 88.6% for correctly detecting positive emotion and 85.8% for correctly detecting negative emotion was achieved using a Receiver Operating Curve (ROC) obtained from 50 simulations. If we were to take the average of these values, we find that the method achieved an average recognition accuracy of 87.2% for both classes of emotion. Ignoring any discrepancies due to differences in testing procedures, we note that the overall accuracy achieved is in the same region (>85%) as that of our 2 class test results without neutral emotion. However, the authors proposed method is only able to discern positive and negative emotion from panic/evacuation situations; which, depending on how the emotion is defined, may not be a true reflection of negative emotion. While this method is limited to panic and evacuation events, our proposed method can be implemented during multiple types of events for the detection of multiple types of emotion.

5   Conclusion

In this paper, we confirmed, via extensive testing on a novel Crowd Emotion dataset with ground-truth emotion, that our proposed Crowd Monitoring algorithm is able to correctly classify a crowd emotion with multiple classes. We found that by excluding neutral emotion and grouping emotions to form a reduced class set, high recognition accuracies were able to be achieved. When testing the performance of our proposed method, it was shown that real-time application is possible. In a comparison with existing methods of Crowd Monitoring in current literature, we found that our proposed algorithm offers a viable alternative to existing techniques. In future work, an improved method of GLTP [15] may be used to further enhance accuracy and efficiency of the algorithm. Implementing

a multiple array camera setup to track faces in 3-Dimensional space will also help to alleviate current limitations with facial obscurities in densely populated crowds.

References

1. Ahmed, F., Hossain, E.: Automated facial expression recognition using gradient-based ternary texture patterns. Chin. J. Eng. **2013**, 1–8 (2013)
2. Ali, S., Shah, M.: A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–6 (2007)
3. Andrade, E.L., Blunsden, S., Fisher, R.B.: Modelling crowd scenes for event detection. In: 18th International Conference on Pattern Recognition, pp. 175–178 (2006)
4. Baig, M.W., Barakova, E.I., Marcenaro, L., Rauterberg, M., Regazzoni, C.S.: Crowd emotion detection using dynamic probabilistic models. In: del Pobil, A.P., Chinellato, E., Martinez-Martin, E., Hallam, J., Cervera, E., Morales, A. (eds.) SAB 2014. LNCS (LNAI), vol. 8575, pp. 328–337. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-08864-8_32
5. Barry, E.: Protests in Moldova Explode, with a Call to Arms on Twitter. The New York Times, p. A1, 7 April 2009
6. Boghossian, B.A., Velastin, S.A.: Motion-based machine vision techniques for the management of large crowds. In: The 6th IEEE International Conference on Electronics, Circuits and Systems, vol. 2, pp. 961–964 (1999)
7. Brostow, G.J., Cipolla, R.: Unsupervised Bayesian detection of independent motion in crowds. In: Computer Society Conference on Computer Vision and Pattern Recognition, pp. 594–601 (2006)
8. Cheriyadat, A.M., Radke, R.: Detecting dominant motions in dense crowds. IEEE J. Sel. Topics Sig. Process. **2**(4), 568–581 (2008)
9. Cikara, M., Jenkins, A.C., Dufour, N., Saxe, R.: Reduced self-referential neural response during intergroup competition predicts competitor harm. NeuroImage **96**, 36–43 (2014)
10. Davies, A.C., Yin, J.H., Velastin, S.A.: Crowd monitoring using image processing. Electron. Commun. Eng. J. **7**(1), 37–47 (1995)
11. Dhall, A.: Context based facial expression analysis in the wild. In: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII), September 2013
12. Ekman, P.: Strong evidence for universals in facial expressions: a reply to russell's mistaken critique. Psychol. Bull. **115**(2), 268–287 (1994)
13. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. **55**, 119–139 (1997)
14. Hassner, T., Itcher, Y., Kliper-Gross, O.: Violent flows: real-time detection of violent crowd behavior. In: 3rd IEEE International Workshop on Socially Intelligent Surveillance and Monitoring (SISM) at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Rhode Island, June 2012
15. Holder, R.P., Tapamo, J.R.: Improved gradient local ternary patterns for facial expression recognition. EURASIP J. Image Video Process. **2017**(1), 42 (2017). https://doi.org/10.1186/s13640-017-0190-5
16. Hsu, C.W., Lin, C.J.: A comparison on methods for multiclass support vector machines. IEEE Trans. Neural Netw. **13**(2), 415–425 (2002)

17. Khryashchev, V., Ganin, A., Golubev, M., Shmaglit, L.: Audience analysis system on the basis of face detection, tracking and classification techniques. In: Proceedings of the International MultiConference of Engineers and Computer Scientists, vol. 1, Hong Kong, March 2013

18. Kumari, J., Rajesh, R., Pooja, K.M.: Facial expression recognition: a survey. In: Second International Symposium on Computer Vision and the Internet, vol. 58, pp. 486–491 (2015)

19. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The extended cohn-kanade dataset (ck+): a complete expression dataset for action unit and emotion-specified expression. In: Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis, San Francisco, CA, USA, pp. 94–101. IEEE (2010)

20. Mahadevan, V., Li, W., Bhalodia, V., Vasconcelos, N.: Anomaly detection in crowded scenes. In: IEEE Computer Vision and Pattern Recognition (2010)

21. Mahto, S., Yadav, Y.: A survey on various facial expression recognition techniques. Int. J. Adv. Res. Electr. Electron. Instrum. Eng. **3**(11), 13028–13031 (2014)

22. Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model. In: IEEE Computer Vision and Pattern Recognition (2009)

23. Mishra, S., Dhole, A.: A survey on facial expression recognition techniques. Int. J. Sci. Res. (IJSR) **4**(4), 1247–1250 (2015)

24. Rabaud, V., Belongie, S.: Counting crowded moving objects. In: Computer Society Conference on Computer Vision and Pattern Recognition, pp. 705–711 (2006)

25. Rabiee, H.R., Haddadnia, J., Mousavi, H., Nabi, M., Murino, V., Sebe, N.: Emotion-based crowd representation for abnormality detection, pp. 1–7. CoRR abs/1607.07646 (2016)

26. Solmaz, B., Moore, B.E., Shah, M.: Identifying behaviors in crowd scenes using stability analysis for dynamical systems. IEEE Trans. Pattern Anal. Mach. Intell. **34**(10), 2064–2070 (2012)

27. Tarjan, R.E.: Minimum spanning trees. In: Data Structures and Network Algorithms, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 44, chap. 6, pp. 72–77. Bell Laboratories, Murray Hill (1983)

28. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, HI, USA, pp. 511–518. IEEE (2001)

29. Wang, X., Ma, X., Grimson, W.E.L.: Unsupervised activity perception in crowded and complicated scenes using hierarchical Bayesian models. IEEE Trans. Pattern Anal. Mach. Intell. **31**(3), 539–555 (2009)

30. Wu, X., Liang, G., Lee, K.K., Xu, Y.: Crowd density estimation using texture analysis and learning, pp. 214–219 (2006)

31. Zhan, B., Monekosso, D.N., Remagnino, P., Velastin, S.A., Xu, L.Q.: Crowd analysis: a survey. Mach. Vis. Appl. **19**(5), 345–357 (2008)

32. Zhang, D., Tong, C., Lu, Y., Liu, Z.: Dominant motions detection in dense crowds based on particle video. Int. J. Digit. Content Technol. Appl. (JDCTA) **6**(10), 294–301 (2012)