



Motion pattern-based crowd scene classification using histogram of angular deviations of trajectories

Abhilash K. Pai¹ · Prahaladh Chandrahasan² · U. Raghavendra³ · A. K. Karunakar¹

Accepted: 6 November 2021 / Published online: 10 January 2022
© The Author(s) 2021

Abstract

Automated crowd behaviour analysis and monitoring is a challenging task due to the unpredictable nature of the crowd within a particular scene and across different scenes. The prior knowledge of the type of scene under consideration is a crucial mid-level information, which could be utilized to develop robust crowd behaviour analysis systems. In this paper, we propose an approach to automatically detect the type of a crowded scene based on the global motion patterns of the objects within the scene. Three different types of scenes whose global motion pattern characteristics vary from uniform to non-uniform are considered in this work, namely structured, semi-structured, and unstructured scenes, respectively. To capture the global motion pattern characteristics of an input crowd scene, we first extract the motion information in the form of trajectories using a key-point tracker and then compute the average angular orientation feature of each trajectory. This paper utilizes these angular features to introduce a novel feature vector, termed as Histogram of Angular Deviations (HAD), which depicts the distribution of the pair-wise angular deviation values for each trajectory vector. Since angular deviation information is resistant to changes in scene perspectives, we consider it as a key feature for distinguishing the scene types. To evaluate the effectiveness of the proposed HAD-based feature vector in classifying the crowded scenes, we build a crowd scene classification model by training the classical machine learning algorithms on the publicly available Collective Motion Database. The experimental results demonstrate the superior crowd classification performance of the proposed approach as compared to the existing methods. In addition to this, we propose a technique based on quantizing the angular deviation values to reduce the feature dimension and subsequently introduce a novel crowd scene structuredness index to quantify the structuredness of an input crowded scene based on its HAD.

Keywords Crowd behaviour analysis · Crowd collectiveness · Crowd scene analysis · Machine learning · Video surveillance

1 Introduction

The rapid advances in technology combined with the continuous growth in the human population have increased the need to develop efficient automated video surveillance-based technologies. As a result, automated crowd video surveillance has

become a popular research area with the motive of ensuring crowd safety. Crowd behaviour analysis [1–3], crowd density estimation/crowd counting [4–6], crowd anomaly detection [7–9], and group detection [10–12] are some of the widely researched areas within the crowd video surveillance domain. In most of these research areas, the performance of the proposed approaches is primarily dependent on the nature and type of the crowded scene [13]. In other words, an approach that performs well for one scene does not guarantee the same performance for a different scenario, especially when the scene dynamics changes.

Analysing the pattern of object motion (people/traffic) within a given crowded scene is one of the effective methods to understand the changes in scene dynamics. According to [14] and [15], crowded scenes can be categorized into *structured* and *unstructured* based on the motion patterns of objects within the scene. A structured scene consists of

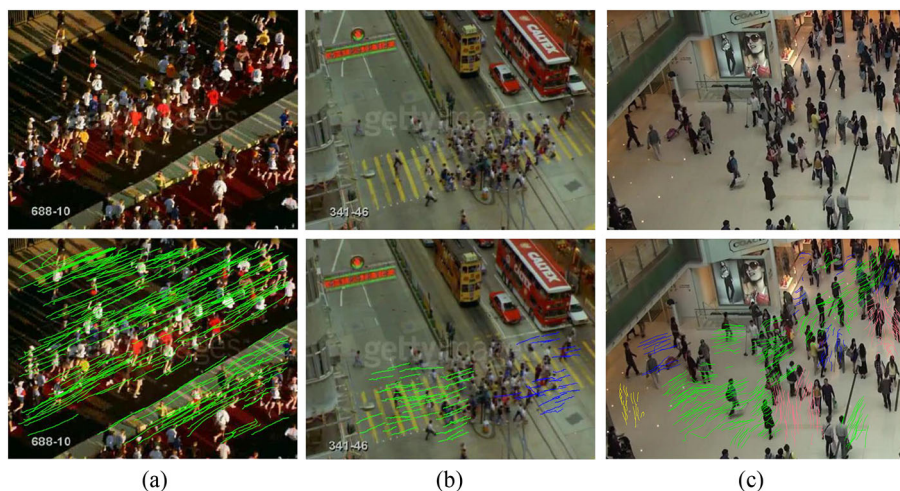
✉ A. K. Karunakar
karunakar.ak@manipal.edu

¹ Department of Computer Applications, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India

² Department of Information & Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India

³ Department of Instrumentation & Control Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India

Fig. 1 Examples of : **a** structured scene and its motion patterns, **b** semi-structured scene and its motion patterns, and **c** unstructured scene and its motion patterns. Note that, the motion patterns are represented in the form of trajectories, with each colour representing different motion patterns (best viewed in colour)



uniform spatio-temporal motion patterns generated by coherently moving objects across the entire scene (Fig. 1a). In other words, in a structured scene, each spatial location contains the same motion pattern, and the direction of motion remains constant for most of the time. In contrast, non-uniform spatio-temporal motion patterns generated by random or chaotically moving objects with unpredictable and frequently varying motion directions make up an unstructured scene (Fig. 1c). However, the scenes with motion patterns that are neither uniform nor chaotic are called *semi-structured* (Fig. 1b).

Identification of the type of a crowded scene is a crucial mid-level information about the scene under consideration. This information helps in the development of an efficient crowd behaviour analysis model. Additionally, this prior knowledge could be applied for crowd monitoring within a particular scene by re-assessing the scene type at particular intervals of time to keep track of the stable-unstable changes in state of the crowd. This paper proposes an approach to classify a given crowded scene into either structured or semi-structured or unstructured based on the motion patterns represented in the form of trajectories. The proposed approach is an extension of our previous work [11] related to crowd motion pattern segmentation using spatio-angular features of the trajectories and an improvised density-based clustering algorithm. Compared to [11], the proposed approach utilizes only the angular features obtained from the trajectories (computed using the gKLT tracker [16]) to compute pair-wise angular deviations between the trajectories. In this work, we additionally compute the histogram of the angular deviations (HAD) which depicts the global motion structure of a scene. To evaluate the HAD-feature's ability, to classify a given scene into either structured, semi-structured and unstructured, we use the publicly available Collective Motion Database to train different classifiers and compare our classification model with the state-of-the-art crowd scene classification approaches which are based on the collec-

tiveness measure. Furthermore, we perform experiments on reducing the original feature dimension by quantizing the angular deviation values into different levels. Finally, using the proposed HAD-based feature vector and a reference histogram for a structured scene, we introduce a measure to quantify the structuredness of a given input scene. The following are the contributions of the proposed work: (i) a novel HAD-based feature vector combined with a robust classifier for efficient crowd scene classification, (ii) an effective quantization-based feature reduction technique for the proposed HAD-feature vector, and (iii) a novel crowd scene structuredness index to quantify the structuredness of a given scene based on its HAD.

2 Related works

While numerous works have been done on motion pattern-based crowd analysis [3,12,17–20], only a few of them focus on classifying a scene into the aforementioned three categories (structured, semi-structured, and unstructured). Among them, Zhou et al. [16,21] introduced a descriptor to quantify a crowded scene based on its 'collectiveness', which is defined as "the degree of individuals acting as a union in collective motion". Three types of collectiveness measures were introduced in their work, namely high, medium and low collectiveness. From their definition, a high degree of collectiveness is a typical characteristics of a structured scene, low degree of collectiveness is seen in unstructured scenes, and semi-structured scenes have medium degree of collectiveness. Since the concepts of high, medium and low collectiveness matches the characteristics of a structured, semi-structured and un-structured crowded scenes, respectively, we compare our work with the collectiveness-based crowd scene classification approaches in the literature. In Zhou et al.'s approach [16,21], a k-Nearest Neighbour

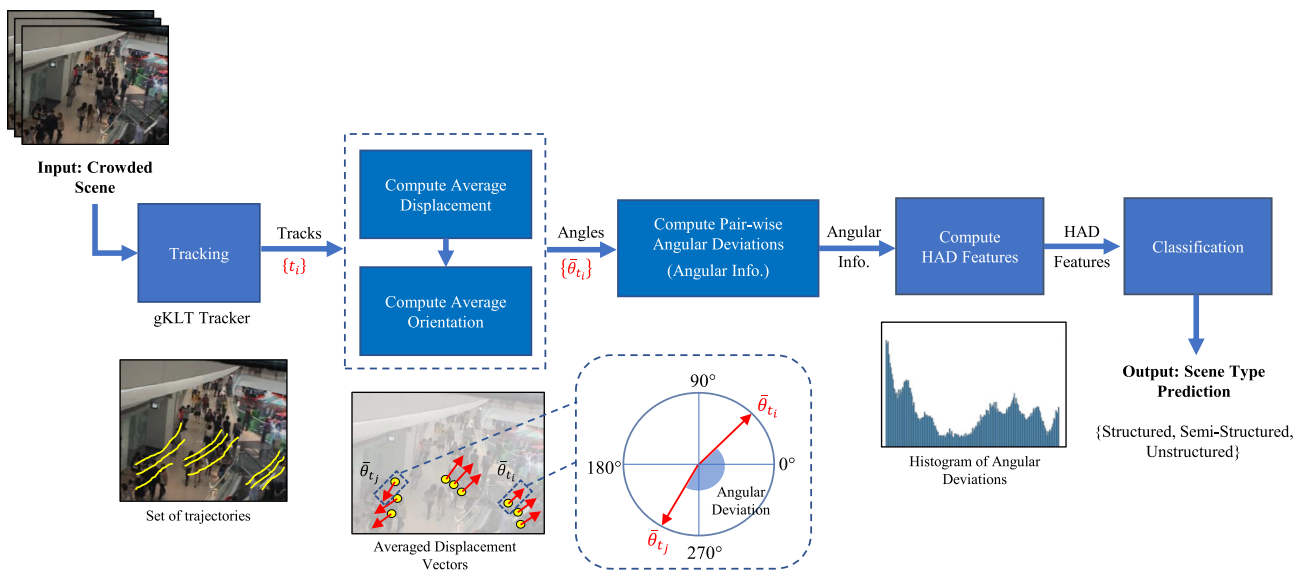


Fig. 2 Block diagram of the proposed approach, where a generalized KLT (gKLT) Tracker [16] is used to extract a set of n trajectories (represented as $\{t_i\}$, where $i = 1 : n$) from the input video. Average angular orientation features (represented as $\{\bar{\theta}_{t_i}\}$) computed from the trajectories is then used to compute angular deviation information between each pair of average angular features $\bar{\theta}_i$ and $\bar{\theta}_j$. The histogram of angular-deviation (HAD) information is subsequently used to train a classifier to predict the scene type

(k-NN) graph is initially constructed with the edge-weights representing the velocity correlation between trajectory vectors. To capture the global behaviour between the trajectory points, the concept of path-based similarity within the weighted graph is employed. Finally, the crowd collectiveness descriptor is computed by aggregating the individual path-based similarities between the nodes in the graph. As a part of this work, Zhou et al. introduced the Collective Motion Dataset (refer Sect. 4.1) for validating the performance of their descriptor. In a similar work, Ren et al. [22] refined the technique for aggregating the topological path-based similarity (introduced in [16]) by using an exponent generating function to produce an improvised collectiveness measure. In another work, Shao et al. [23,24] used a Markov chain-based approach and proposed a Collective Transition (CT) Prior to model the crowd group behaviour. The CT-prior is then used to define group-level collectiveness of the crowd. In contrast, Li et al. [25] used a point selection strategy to refine the features points which are tracked, followed by a manifold ranking approach to compute the crowd collectiveness. In an extended work, Li et al. [10,26] modelled the motion intention of individuals in a scene by proposing an intention-aware model which is combined with a manifold ranking strategy to compute the collectiveness of the crowd. Recently, Roy et al. [27] proposed an approach to classify a given crowded scene into structured, semi-structured and unstructured based on the definitions presented in [15].

Roy et al. [27] used the direction property of the crowd and divided each frame into non-overlapping blocks to compute angle/orientation histograms for each block from the trajectory data. Before combining all the block-level orientation histograms, a Gaussian averaging and an angular value-based quantization approach are used to refine each of them.

Most of the aforementioned approaches measure frame-by-frame collectiveness value and subsequently average them (for all the frames) to compute the total collectiveness of a given video. Such frame-by-frame approaches lead to huge variations in the measured collectiveness because of the continuous change in the motion patterns of the trajectory key-points between frames. Also, these approaches are heavily dependent on their model initialisation parameters and are computationally complex. In contrast, we propose a multi-frame approach by simply averaging the trajectory data for fixed set of frames, which not only enables to capture the history of motion but also produces a more stable feature vector (to quantify crowd collectiveness) as compared to the frame-by-frame approaches. Furthermore, even though our approach is similar to [27] in using histogram based on the direction property of the crowd, the essential difference lies with the use of a more robust and distinctive histogram of angular deviation instead of histogram of angular orientation (explained in Sect. 3). Finally, our approach is parameter-free and is computationally inexpensive.

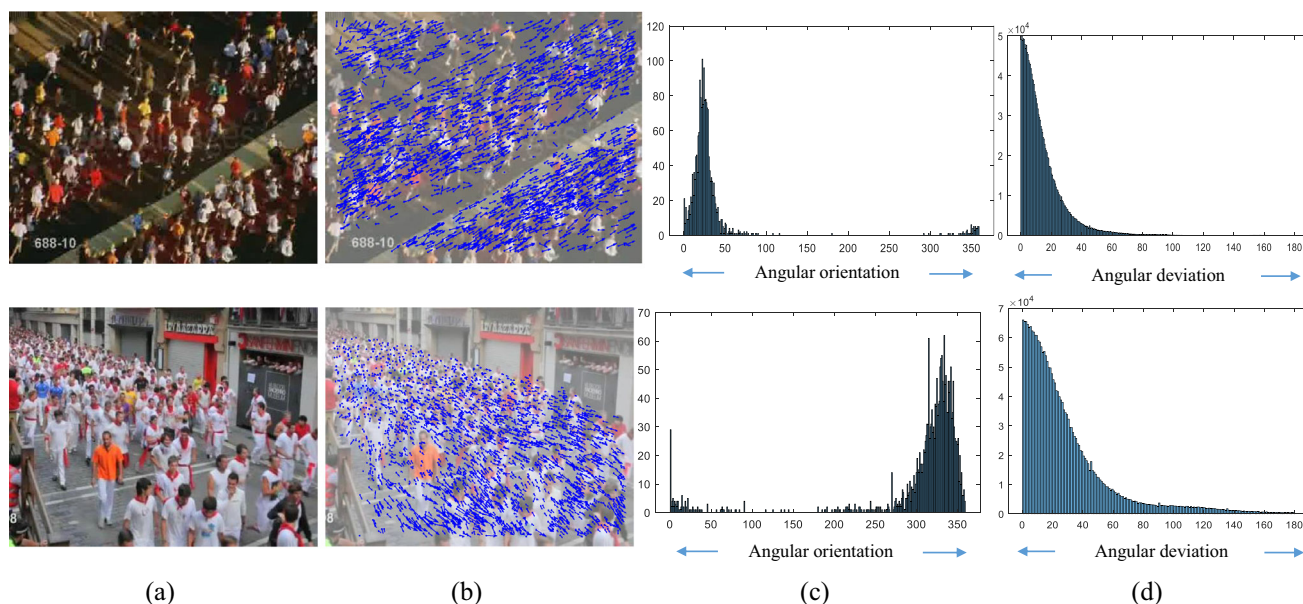


Fig. 3 Histogram of Angular Orientations versus Histogram of Angular Deviations. Each row represents data related two different structured scenes: **a** Frames from each of the two scenes, **b** Averaged displacement vectors overlaid on the frame, **c** Histogram of angular orientations, **d** Histogram of angular deviations

3 Crowd scene classification using Histogram of Angular Deviations

According to [16,28–30], from the macroscopic point-of-view, the moving crowd has a high degree of collective or structured behaviour if majority of the participants in the crowd move together uniformly in the same direction. Based on this property of the crowd, we propose an approach to classify a given crowded video into either structured, semi-structured or unstructured by computing the angular orientation information of the moving crowd. The proposed approach is detailed in the following paragraphs according to the block diagram shown in Fig. 2.

Extract Trajectories(Tracking): Given an input crowd video, the first step is to capture the motion of the crowd. This is done by using an accurate and computationally efficient generalized KLT (gKLT) tracker introduced by Zhou et al. [16], which detects and tracks key-points (corner features) between consecutive frames of the video. Factors like occlusion and illumination variations can produce noisy trajectories with very short length or instances of zero-displacement values for the major part of a trajectory. Such noisy trajectories are discarded using pre-defined thresholds (determined empirically).

Compute Average Angular Orientation: A trajectory is a set of 2D coordinate points that depict a key-point (belonging to an object in the scene) movement across a set of consecutive frames. An effective approach to capture the direction of the moving crowd from the trajectory data is to first compute the

average of frame-by-frame displacements for each trajectory. The average angular orientation $\bar{\theta}_i$ for each trajectory is then obtained from the already computed average displacement vector by projecting it onto a unit vector in the horizontal direction (along x-axis), as demonstrated in [11].

Compute Histogram of Angular Deviations (HAD): For each trajectory, the average angular orientation value signifies the history of the trajectory movement for a period of time. If all the averaged trajectory vectors in the scene approximately point towards the same direction, then the scene is said to be structured. If the trajectory vectors are scattered in different directions, then it characterizes an unstructured scene. Therefore, analysing the distribution of orientation values or the histogram of angular orientation gives a clear picture of the global movement characteristics of objects within a scene. However, as shown in Fig. 3c, even though the histogram of angular orientations for two different structured scenes clearly characterizes the structured behaviour, it produces peaks at different locations of the histogram. Therefore, the values of the histogram of angular orientations cannot be considered as a distinguishing feature in its original form. Hence, in this work, we compute a robust histogram of angular deviations instead. The angular deviation ($\Delta\bar{\theta}_{t_{ij}}$) between two averaged trajectory vectors $\bar{\theta}_{t_i}$ and $\bar{\theta}_{t_j}$ is defined as follows [11]:

$$\Delta\bar{\theta}_{t_{ij}} = \min(|\bar{\theta}_{t_i} - \bar{\theta}_{t_j}|, 360 - |\bar{\theta}_{t_i} - \bar{\theta}_{t_j}|) \quad (1)$$

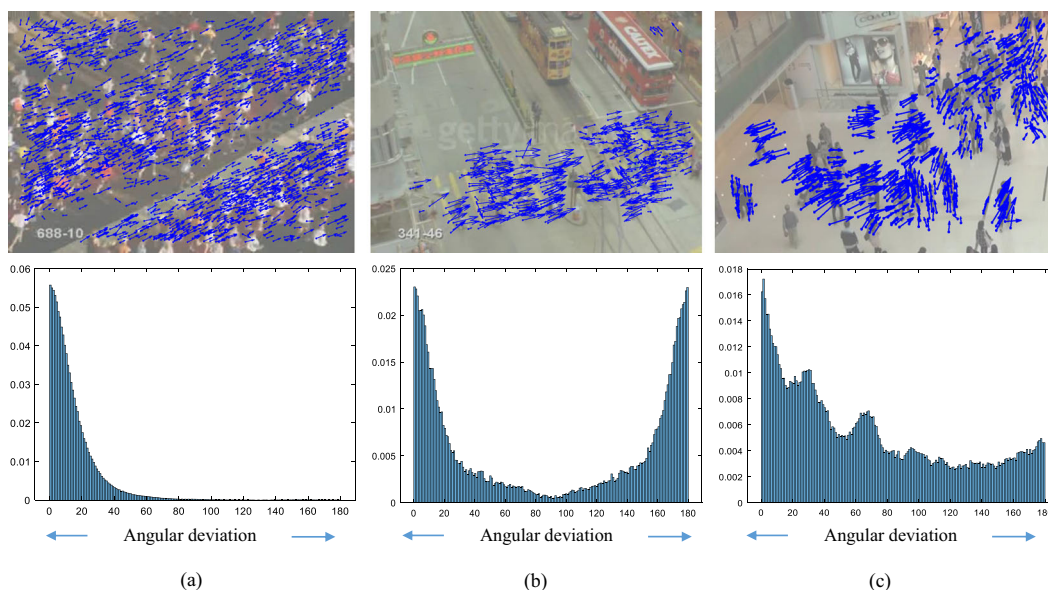


Fig. 4 **a** Structured scene and its HAD, **b** Semi-structured scene and its HAD, and **c** Unstructured scene and its HAD

where the range of $\bar{\theta}_{t_{ij}}$ is between 0° and 180° (inclusive of both). Computing the histogram of pair-wise angular deviation values (for each trajectory vector with every other trajectory vectors) provides a global/scene-level picture of correlation of the trajectory vectors in terms of angles. If the histogram is peaked close to the angular deviation value of 0° , it means that the majority of trajectory vectors in the scene move in the same direction (structured scene). Therefore, as shown in Fig. 3d, the HAD produces consistent peaks for different scenes of the same type as compared to the histogram of angular orientations. Given an input scene, we compute the HAD and consider the values of the HAD as a feature vector representing the type of scene. Since the range of $\Delta\bar{\theta}_{t_{ij}}$ is $[0^\circ, 180^\circ]$, the dimension of the HAD-feature vector is 181. The HADs for three different types of scenes, shown in Fig. 4, justifies the discriminative capability of the proposed HAD feature vector.

Classification: The HAD-features are used to build a supervised machine learning model for classifying a given crowded scene into either of the three categories. For this purpose, we use the Collective Motion Database [16] (refer section 4.1, for more details of the dataset). Since this dataset con-

tains less data (413 videos) and the proposed HAD-feature vector dimension is large, we choose the nonlinear classical machine learning algorithms for our experimentation. Based on exhaustive experimentation over different classical machine learning algorithms, we observe that the proposed HAD feature vector can efficiently discriminate most of the scenes in the Collective Motion Database. The results of this experimentation are presented in Sect. 4, where for the sake of simplicity, we have selected the best-performing classical machine learning algorithms, namely Weighted k-Nearest Neighbours (Weighted k-NN) [31], ν -Support Vector Machines (ν -SVM) [32], and Tree ensembles based on an efficient implementation of gradient boosting called eXtreme Gradient Boosting (XGBoost) [33].

4 Experimental results & discussion

4.1 Dataset

The *Collective Motion Database* (introduced by Zhou et al. [16]) is used for evaluating the effectiveness of the pro-

Table 1 Statistics for the Collective Motion Database

Scene type (Label-ID)	No. of scenes
Structured (2)	90
Semi-structured (1)	107
Unstructured (0)	216
Total	413

Table 2 Comparison of binary classification performances (measured using the performance metrics—Precision (P), Recall (R), and F1-Score) of the proposed approach with the state-of-the-art on the Collective Motion Database (The best score is marked in bold)

Approach	Structured- Unstructured			Structured-Semi-structured			Semi-structured- Unstructured		
	P	R	F1- Score	P	R	F1- Score	P	R	F1- Score
Zhou et al. [16]	0.81	0.58	0.51	0.76	0.57	0.48	0.74	0.47	0.40
Shao et al. [24]	0.88	0.60	0.58	0.79	0.55	0.52	0.73	0.49	0.44
Li et al. [10]	0.92	0.71	0.75	0.87	0.70	0.69	0.83	0.72	0.65
Proposed HAD (Weighted k-NN)	0.94	0.92	0.93	0.87	0.88	0.87	0.86	0.85	0.85
Proposed HAD (ν -SVM)	0.95	0.94	0.95	0.89	0.88	0.88	0.84	0.84	0.84
Proposed HAD (XGBoost)	0.95	0.95	0.95	0.86	0.85	0.85	0.84	0.84	0.84

posed approach. The dataset contains a total of 413 crowd video sequences, with each video containing 100 frames. The dataset also contains ground truth labels for each scene (the labels belong to the set {0, 1, 2}), where ‘0’ refers to scenes with low-collectiveness (unstructured scenes in the context of our work), ‘1’ refers to scenes with medium-collectiveness (semi-structured scenes) and ‘2’ refers to scenes with high-collectiveness (structured scenes). The ground truth labels are generated by major voting the manually decided labels by 10 human subjects, for each scene. Data statistics in Table 1 shows that 52% of the dataset consists of unstructured scenes, while 22% are structured and the rest 26% are semi-structured, which means that the dataset is imbalanced.

4.2 Experimental set-up

For each scene of the Collective Motion Database, a set of 3000 key-points (as per [16]) are detected and tracked for entire duration of the video using the gKLT tracker. The average displacement is computed for each set of 30 consecutive frames. For the ease of evaluation we experimentally choose 30 consecutive frames from the entire trajectory data which best represents the crowd behaviour (as per [24]). The noisy trajectories are then filtered by discarding the trajectories with total length less than 5-frames, and zero-displacement throughout the trajectory’s span (occurring due to tacking/motion estimation errors). After computing the HAD, the count values of the histogram are normalized in the range [0, 1]. Since the dataset is imbalanced, classifier’s performance is evaluated by employing a stratified 10-fold cross-validation with Precision, Recall and F1-Score chosen as the evaluation metrics [34]. The essential hyperparameters (determined empirically based on repeated experiments) used to configure the three classifiers are: (i) Weighted k-NN model uses Bhattacharya distance measure [35] with optimal results obtained for the k-value of 10, (ii) the ν -SVM model uses a Radial Basis Function (RBF)-based kernel with an ν value of 0.3, and (iii) XGBoost-based tree ensemble model

uses 30 trees for each ensemble with maximum depth set to 3 and a learning rate of 0.1.

4.3 Results & discussion

Firstly, we compare the classification performance of the proposed HAD-based approach with the state-of-the-art approaches [16], [24], [10], which reports only binary classification results for the pair-wise combinations of structured vs. structured, structured vs. semi-structured, and semi-structured vs. unstructured scenes. The results shown in Table 2 (results for the existing approaches are obtained from the paper [10]) clearly indicates the superior performance of the proposed HAD-based approach, which is also able to

Table 3 3-Class Classification (Structured, Semi-structured, and Unstructured) performance for the proposed HAD-based approach using the selected classifiers (The best score is marked in bold)

Classifier	Precision	Recall	F1-Score
Weighted k-NN	0.84	0.82	0.82
ν -SVM	0.83	0.80	0.81
XGBoost	0.82	0.80	0.80

Table 4 Confusion matrix for 3-Class Classification (0: Unstructured, 1: Semi-structured, 2: Structured) using the Weighted k-NN classifier

		Actual class			
		0	1	2	
Predicted class	0	199	14	3	0
	1	20	74	13	1
	2	3	10	77	2
		0	1	2	

perform consistently well for each of the selected three classifiers. The state-of-the-art approaches work by modelling complex interactions among the trajectories to generate a collectiveness measure. Since this measure is a single value and is heavily dependent on the scene dynamics and subsequently the model parameters, a small amount of anomaly can adversely effect the collectiveness value, which is a major reason for the low scores for most state-of-the-art approaches.

On the contrary, the proposed HAD-based approach depends only on the accuracy of the tracker (which is applicable to the state-of-the-art approaches as well) and comprises of a multi-valued feature vector where a small amount of anomaly does not create a drastic change in the overall feature vector. It is also observed that the proposed approach is most effective in classifying structured and unstructured scenes. This is because, as demonstrated in Sect. 3, the dif-

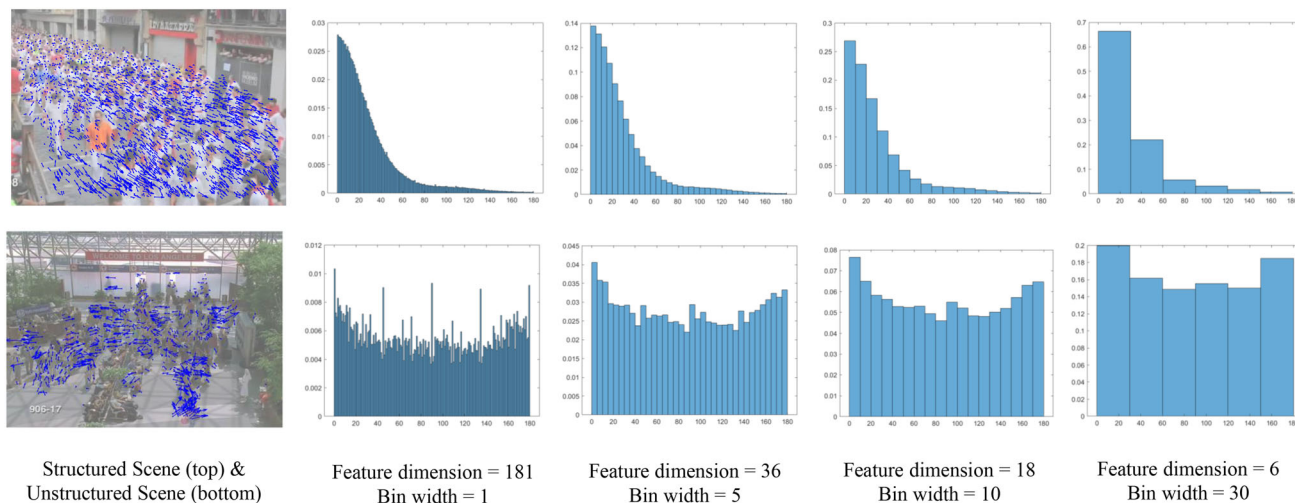


Fig. 5 Visualization of the HADs for each level of quantization of the angular deviation values. Each row represents the HADs for structured (top-row) and unstructured scenes (bottom-row). From left to right: Frames from the input scene with trajectory vectors, the original HAD

with feature vector dimension 181 (bin width = 1), and some of the quantized HADs with feature vector dimensions: 36 (bin width = 5), 18 (bin width = 10), and 6 (bin width = 30)

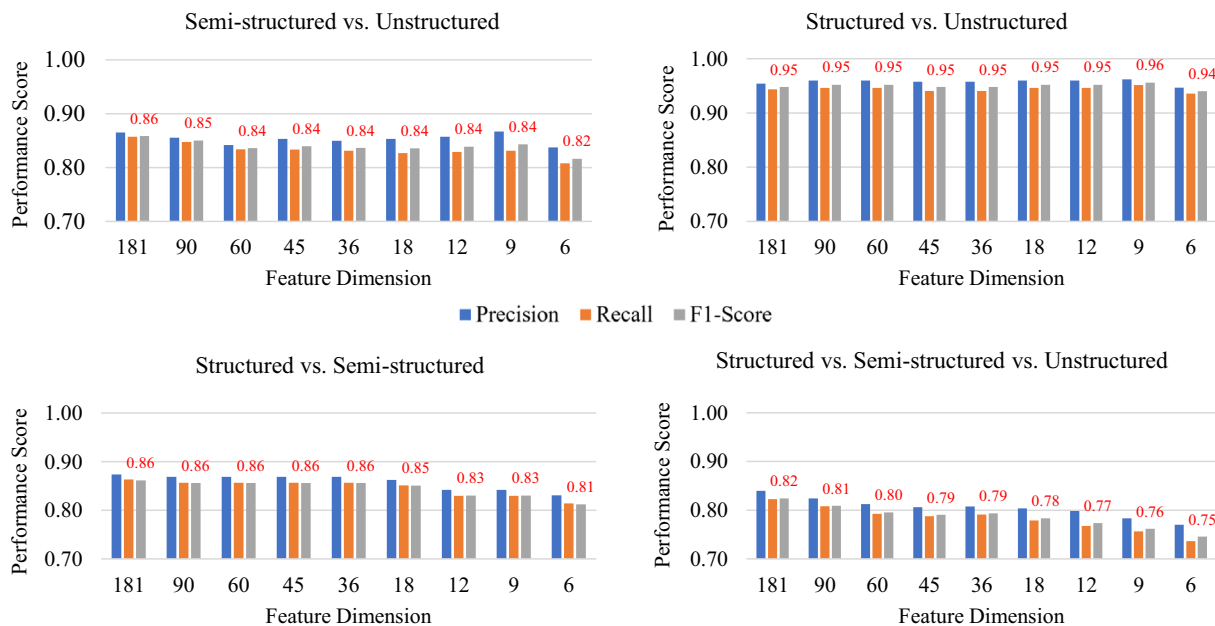


Fig. 6 Binary and 3-class classification performances of the proposed approach for varying feature dimension using Weighted k-NN classifier, where the feature dimension is varied by quantizing the angular deviation values (the F1-scores for all the approaches are highlighted in red-font for easier interpretation)

ferent possible values for angular deviations for a structured scene is very less and is close to 0° . Whereas, an unstructured scene contains objects moving in different directions due to which a variety of values are possible for the angular deviation. However, there are instances of several misclassification in the case of structured vs. semi-structured and semi-structured vs. unstructured classifications, due to the close resemblance of the semi-structured scenes with both structured and unstructured scenes. This is also evident from Tables 3 and 4, where we list out the three-class (structured, semi-structured, unstructured) classification performance of the proposed approach for the selected classifiers. Furthermore, since Roy et al. [27] uses a different proportion of scenes for each class (as compared to Table 1) and their binaries are not available, we are unable to quantitatively compare their results. Nonetheless, the dimension of the proposed HAD-based feature vector is 20-times lesser than the feature vector proposed by Roy et al. [27]. More importantly, the proposed HAD-based approach uses the angular deviation measure, which is not only globally consistent within a scene but also consistent across different scenes (Figs. 3 and 4), as compared to the angular orientation-based feature used in [27].

Secondly, we examine the effect of reducing the feature dimension by quantizing the angular deviation values. For this purpose, we choose the weighted k-NN classifier based on the Bhattacharya Distance measure [35], since it generalizes well to reduction in feature dimension. The original angular deviation values in the range $[0^\circ, 180^\circ]$ (having 181 levels with $bin\ width = 1$) are quantized into different levels, viz. 90, 60, 45, 36, 18, 12, 9, and 6 with each level having a histogram bin width of 2, 3, 4, 5, 10, 15, 20 and 30, respectively (the last bin will have $bin\ width = bin\ width + 1$). As a result of this quantization, adjacent angular deviation values within a particular range (defined by the bin width corresponding the quantization level) contribute to a single histogram bin count. The visualization of the changes in the HAD structure after the quantization operation (for different levels) can be realised from Fig. 5. It is observed that the HADs, for all levels of quantization, retains the same global structure. Figure 6 shows the performance metrics for each of the reduced feature vector applied over the Collective Motion database, where it is observed that increasing the bin width does not drastically decrease the classification performance. More importantly, the reduced feature vectors can distinguish the structured and unstructured scenes effectively. This is primarily because of the nature of the two scenes. In structured scenes, the pair-wise angular deviation between the majority of trajectory vectors is minimal, resulting in a peak close to the 0° bin in the histogram. Whereas, in unstructured scenes, the pair-wise angular deviation between the trajectory vectors is not minimal and can take up any values, resulting in the values distributed across the histogram with no clear peaks.

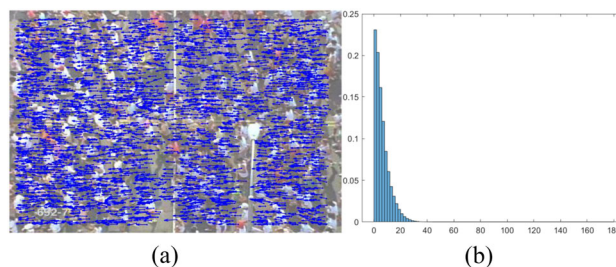


Fig. 7 **a** The reference structured scene chosen from the Collective Motion Dataset (scene name: “startRunning5”), and **b** its HAD with feature dimension = 90, bin width = 2

Hence, increasing the bin width does not effect the major trend in HAD and thus the capability for discriminating the structured and unstructured scenes. However, it can be also be observed that decreasing the feature dimension limits the capability of the classifier in discriminating subtle variations in the scenes, especially in the case of some semi-structured scenes whose HADs are either similar to that of unstructured or structured scenes. Since, reduction in feature dimension results in faster computations, we focus on choosing a feature dimension from Fig. 6 which does not significantly reduce the classification performance. Based on the data in Fig. 6, we make the following two conclusions: (i) the classification performance of the reduced feature vector with dimension equal to 90 ($binwidth = 2$) is equivalent to overall performance of the original feature vector (for 2-class and 3-class classification). Thus, the HAD with feature dimension equal to 90 is considered as an optimal choice for crowd classification. (ii) the HAD with feature dimension equal to 9 can be used for structured-unstructured scene classification.

4.4 Crowd scene structuredness index

Based on the Collective motion dataset and its HAD data (for each scene), we empirically define a measure, termed as the crowd scene structuredness index, to quantify the structuredness for a given scene (in the scale of 0 to 1). For this purpose, we consider a reference histogram to be one of the scenes (shown in Fig. 7) within the Collective motion dataset which is closest to an ideal structured scene. Now, for a given input scene, we compute the HAD (here, we use the HAD with feature vector dimension 90) and then, compare it with the reference histogram using the Bhattacharya coefficient measure as shown in Eq. 2.

$$\phi(r, q) = \sum_{x \in B} \sqrt{r(x)q(x)} \quad (2)$$

where, the crowd scene structuredness index is represented using the function ϕ , which quantifies the amount of struc-

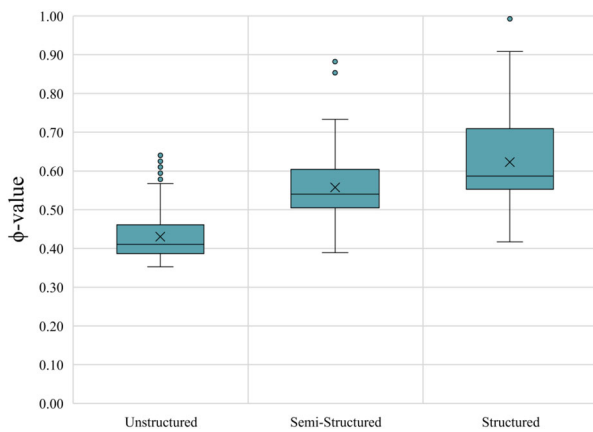


Fig. 8 Box plot of the crowd scene structuredness index (ϕ -value) for structured, semi-structured and unstructured scene categories (the ϕ -value is computed for all the scenes of the Collective motion dataset and the box plots are computed category-wise)

turedness of the computed HAD (denoted as q) when compared to the reference HAD (denoted as r), and B denotes the set of all bins in the histogram. The entire term in the right hand side of Eq. 2 represents the Bhattacharya coefficient, a part of the Bhattacharya distance metric, which is widely used to measure the similarity of two distributions. The statistical details of this experiment are shown in Fig. 8, which clearly indicate the separability between the structured and unstructured classes. The following can be concluded from Fig. 8: (i) the scenes with ϕ -value greater than 0.55 have structured motion patterns (greater the ϕ -value, more the structuredness), (ii) the scenes with ϕ -value less than or equal to 0.55 have unstructured motion patterns (lower the ϕ -value, greater the unstructuredness), and (iii) the scenes with values between 0.40 and 0.65 are semi-structured with values greater than 0.55 indicating a small amount of structuredness within the scene and values less than or equal to 0.55 indicating a small amount of unstructuredness within the scene.

5 Conclusion

In this paper, a feature vector based on the histogram of angular deviations (HAD) of averaged trajectory vectors was proposed to classify a given crowded scene into structured, semi-structured and unstructured, based on global motion patterns. The proposed HAD-based feature vector is composed of values which depict the count of each possible angular deviation value between 0° and 180° . Since a structured scene contains objects moving uniformly in the same direction, the pair-wise angular deviations between the majority of trajectory vectors will be minimal. This creates a well distinguishable peak in the HAD close to the 0° -bin.

Whereas, for an unstructured scene, the angular deviations values are distributed across various bins of HAD, due to the motion of the objects in different directions. Based on this notion, the experiments performed on the publicly available Collective Motion Database using classical machine learning algorithms prove the robustness of the proposed HAD-feature in distinguishing different scene types, specifically structured and unstructured scenes. Furthermore, the proposed approach outperforms the state-of-the-art approaches in binary classification of different combinations of the various scene types. The experiments conducted on quantizing the angular deviation values to reduce the feature dimension proved that the reduced feature vector with dimension equal to 90 (bin width equal to 2) performs as good as the original feature vector of dimension equal to 181, making it an optimal feature dimension for classifying crowded scenes based on motion patterns. Finally, based on comparing the proposed HAD with a reference HAD which depicts an ideal structured scene, we define a crowd scene structuredness index (based on Bhattacharya coefficient) which quantifies the amount of structuredness in a given scene. As a future work, we intend to improvise the computation of HAD by introducing a penalty function that penalizes unstructured behaviour, thereby improving the 3-class classification performance and increasing the separation between the three classes. Furthermore, in order to distinguish between all the three types of crowded scenes even better, we intend to explore about the possibility of including additional features other than the angular deviation feature.

Funding Open access funding provided by Manipal Academy of Higher Education, Manipal.

Declarations

Conflict of interest The authors declare that they have no conflict of interest

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Haghani, M., Sarvi, M.: Crowd behaviour and motion: empirical methods. *Transp. Res. Part B Methodol.* **107**, 253–294 (2018)
- Tripathi, G., Singh, K., Vishwakarma, D.K.: Convolutional neural networks for crowd behaviour analysis: a survey. *Vis. Comput.* **35**(5), 753–776 (2019)
- Farooq, M.U., Saad, M.N.M., Khan, S.D.: (2021) Motion-shape-based deep learning approach for divergence behavior detection in high-density crowd. *Vis. Comput.* pp 1–25
- Gao, G., Gao, J., Liu, Q., Wang, Q., Wang, Y.: (2020) Cnn-based density estimation and crowd counting: A survey. *arXiv preprint arXiv:2003.12783*
- Khan, S.D., Basalamah, S.: Scale and density invariant head detection deep model for crowd counting in pedestrian crowds. *Vis. Comput.* **37**(8), 2127–2137 (2021)
- Sindagi, V.A., Patel, V.M.: A survey of recent advances in cnn-based single image crowd counting and density estimation. *Pattern Recogn. Lett.* **107**, 3–16 (2018)
- Bansod, S.D., Nandedkar, A.V.: Crowd anomaly detection and localization using histogram of magnitude and momentum. *Vis. Comput.* **36**(3), 609–620 (2020)
- Sánchez, F.L., Hupont, I., Tabik, S., Herrera, F.: Revisiting crowd behaviour analysis through deep learning: Taxonomy, anomaly detection, crowd emotions, datasets, opportunities and prospects. *Information Fusion* (2020)
- Afiq, A., Zakariya, M., Saad, M., Nurfarzana, A., Khir, M.H.M., Fadzil, A., Jale, A., Gunawan, W., Izuddin, Z., Faizari, M.: A review on classifying abnormal behavior in crowd scene. *J. Vis. Commun. Image Represent.* **58**, 285–303 (2019)
- Li, X., Chen, M., Wang, Q.: Quantifying and detecting collective motion in crowd scenes. *IEEE Trans. Image Process.* **29**, 5571–5583 (2020)
- Pai, A.K., Karunakar, A.K., Raghavendra, U.: Scene-independent motion pattern segmentation in crowded video scenes using spatio-angular density-based clustering. *IEEE Access* **8**, 145984–145994 (2020). <https://doi.org/10.1109/ACCESS.2020.3015375>
- Lamba, S., Nain, N.: (2019) Segmentation of crowd flow by trajectory clustering in active contours. *Vis. Comput.* pp 1–12
- Zitouni, M.S., Bhaskar, H., Dias, J., Al-Mualla, M.E.: Advances and trends in visual crowd analysis: a systematic survey and evaluation of crowd modelling techniques. *Neurocomputing* **186**, 139–159 (2016)
- Rodriguez, M., Ali, S., Kanade, T.: (2009) Tracking in unstructured crowded scenes. In: 2009 IEEE 12th International Conference on Computer Vision, IEEE, pp 1389–1396
- Li, T., Chang, H., Wang, M., Ni, B., Hong, R., Yan, S.: Crowded scene analysis: a survey. *IEEE Trans. Circuits Syst. Video Technol.* **25**(3), 367–386 (2014)
- Zhou, B., Tang, X., Zhang, H., Wang, X.: Measuring crowd collectiveness. *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(8), 1586–1599 (2014)
- Lu, W., Wei, X., Xing, W., Liu, W.: Trajectory-based motion pattern analysis of crowds. *Neurocomputing* **247**, 213–223 (2017)
- Wu, S., Yang, H., Zheng, S., Su, H., Zhou, Q., Lu, X.: Motion sketch based crowd video retrieval. *Multimed. Tools Appl.* **76**(19), 20167–20195 (2017)
- Zhao, W., Zhang, Z., Huang, K.: Gestalt laws based tracklets analysis for human crowd understanding. *Pattern Recogn.* **75**, 112–127 (2018)
- Cheng, Y., Yang, H., Chen, L.: (2019) An online crowd semantic segmentation method based on reinforcement learning. In: 2019 IEEE International Conference on Image Processing (ICIP), IEEE, pp 2429–2433
- Zhou, B., Tang, X., Wang, X.: (2013) Measuring crowd collectiveness. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3049–3056
- Ren, W.Y., Li, S.H., Guo, Q., Li, G.H., Zhang, J.: (2015) Agglomerative clustering and collectiveness measure via exponent generating function. *arXiv preprint arXiv:1507.08571*
- Shao, J., Change Loy, C., Wang, X.: (2014) Scene-independent group profiling in crowd. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 2219–2226
- Shao, J., Loy, C.C., Wang, X.: Learning scene-independent group descriptors for crowd understanding. *IEEE Trans. Circuits Syst. Video Technol.* **27**(6), 1290–1303 (2016)
- Li, X., Chen, M., Wang, Q.: Measuring collectiveness via refined topological similarity. *ACM Trans. Multimed. Comput. Commun. Appl. TOMM* **12**(2), 1–22 (2016)
- Wang, Q., Chen, M., Li, X.: (2017) Quantifying and detecting collective motion by manifold learning. In: Proceedings of the AAAI Conference on Artificial Intelligence
- Roy, A., Biswas, N., Saha, S.K., Chanda, B.: (2019) Classification of moving crowd based on motion pattern. In: 2019 IEEE Region 10 Symposium (TENSYP), IEEE, pp 102–107
- Le Bon, G.: (1897) *The crowd: A study of the popular mind*. Macmillan
- Raafat, R.M., Chater, N., Frith, C.: Herding in humans. *Trends Cogn. Sci.* **13**(10), 420–428 (2009)
- Vicsek, T., Zafeiris, A.: Collective motion. *Phys. Rep.* **517**(3–4), 71–140 (2012)
- Dudani, S.A.: The distance-weighted k-nearest-neighbor rule. *IEEE Trans. Syst. Man Cybern. SMC* **6**(4), 325–327 (1976)
- Schölkopf, B., Smola, A.J., Williamson, R.C., Bartlett, P.L.: New support vector algorithms. *Neural Comput.* **12**(5), 1207–1245 (2000)
- Chen, T., Guestrin, C.: (2016) Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp 785–794
- Olson, D.L., Delen, D.: *Advanced Data Mining Techniques*. Springer Science & Business Media, Berlin (2008)
- Kailath, T.: The divergence and bhattacharyya distance measures in signal selection. *IEEE Trans. Commun. Technol.* **15**(1), 52–60 (1967)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Abhilash K. Pai received the M. Tech and B. Tech degrees from the Cochin University of Science and Technology, Cochin, Kerala in 2014 and 2010, respectively. He is currently a Ph.D. research scholar at the Department of Computer Applications, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India. His research interests include computer vision and machine learning.



Prahaladh Chandrahasan is currently a Third year B. Tech student at the Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India. His research interests include computer vision, reinforcement learning and machine learning.



U. Raghavendra received the Ph.D. degree from the Manipal Academy of Higher Education, India. He is currently a Faculty with the Department of Instrumentation and Control Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India. He has published several papers in refereed international SCI-IF journals and international conference proceedings. He has a patent to his credit and received an invention award from Intellectual Ventures,

USA, for his innovations in 2014. His current research interests include 3D computer vision, image processing, and medical image analysis.



A. K. Karunakar received the B. Sc degree and M.C.A degree from Karnataka University, Karnataka, India, in 1995 and 1998, respectively, and the Ph. D degree from Manipal Academy of Higher Education, Karnataka, India in 2009. He is currently a Professor and Head, Department of Computer Applications, from Manipal Academy of Higher Education, Karnataka, India. His research interest includes image/video processing and communication, Scalable Video Coding, Media Aware Network Elements, Multi-view Video Coding, Scalable video over Peer-to-Peer Networks, Error resilient and concealment for scalable video, stereo vision, image & video forensics.