# Machine Learning for Dense Crowd Direction Prediction Using Long Short-Term Memory

Abdullah Alajlan
*Computer Science Department*
*University of Idaho*
*Moscow, USA*
*Computer Science Department*
*Technical and Vocational Training Corporation*
*Riyadh, 12613, Saudi Arabia*
*Email: alaj0169@vandals.uidaho.edu*

Alaa Edris
*Computer Science Department*
*University of Idaho*
*Moscow, USA*
*University of Jeddah*
*Jeddah, 23890, Saudi Arabia*
*Email: edri6611@vandals.uidaho.edu*

Frederick Sheldon
*Computer Science Department*
*University of Idaho*
*Moscow, USA*
*Email: sheldon@uidaho.edu*

Terence Soule
*Computer Science Department*
*University of Idaho*
*Moscow, USA*
*Email: tsoule@uidaho.edu*

***Abstract*—The safety of a dense crowd is one of the most important matters for an event's organizers. Therefore, management of the crowd, and noticing any serious issues in advance becomes important. Developing a crowd simulation by using a social force model simulates the behavior of crowds in reality. The prediction of individual agents' behavior in the simulation and how the agents interact with each other can improve the safety of dense crowds. Depending on the success of Recurrent Neural Network(RNN) handling of sequential data, we propose a model that is based on Long Short Term Memory (LSTM) to predict individual agents' next locations. Our proposed approach will be tested two different densities of crowds, structured crowds, and unstructured crowds. In structured crowds, people generally move in one direction and head to the same destination, such as at the Islamic Hajj. In unstructured crowds people move in many different directions and head for different destinations, such as in public town squares.**

***Keywords*—Crowds, Structured Crowded Area, Unstructured Crowded Area, Machine Learning, Recurrent Neural Network, and Long Short-Term Memory, "Short Paper" Symposium on Artificial Intelligence (CSCI-ISAI).**

(a) Example of an unstructured crowd.Image from Ozturk et al. [11].

(b) Example of structured crowd. Image is a screen capture from videos in 2019 from Ministry of Hajj, Kingdom of Saudi Arabia, https://www.haj.gov.sa/en

Figure 1. Examples of two forms of crowds.

## 1. Introduction

People usually follow rules that are taken for granted when they walk in crowds. For instance, in a dense crowd, individuals move to their next locations while avoiding people or obstacles in front of them. Understanding these rules leads us to avoid dangerous situations and maintains the safety and stability of crowds. [12] has shown that there is one type of crowd motion in a structured crowd, such as at the Hindu Kumbh, and another type of crowd motion in an unstructured crowd, such as in subway stations. Most prediction simulations have been done without differentiation between the types of crowds. Our proposed approach is to employ machine learning to understand how people behave in the two different types of crowds, structured, and

978-1-7281-7624-6/20/$31.00 ©2020 IEEE
DOI 10.1109/CSCI51800.2020.00125

unstructured.

Recurrent Neural Networks(RNN) and especially Long Short-Term Memory(LSTM) networks have become a very popular method to understand the sequential nature of their inputs. LSTM has shown promising results in problems with sequential data, such as individuals' trajectories, vehicle motion, handwriting, and speech. [1] presented how to connect LSTM networks for every trajectory in relation to each other, which in turn, allows every LSTM network to share information with close networks. [12] has done experiments that depend on two types of crowds (structured and unstructured crowds) based on the closest people in the cone of vision. Prediction of individuals' movements could be improved if we take into consideration the differences in crowd behavior in each type of crowd. This will make it easier for machine learning to accurately learn each kind of movement. Focusing on the crowd type to acquire data is a form of classification. Our proposed model is an extension of [2], which use a cone of vision to specify the direction of motion based on the closest people; furthermore, employs the LSTM networks technique to monitor the previous cone of vision direction for every individual in order to understand each individual's behavior. Additionally, the crowd results will be divided into two categories (structured, and unstructured crowds)

We present related work in the following section. This will be followed by the Methodology and Datasets sections. Finally, we will present the conclusion of our proposed model.

## 2. Background

Based on the past locations, Alahi et al. [1] used the LSTM model to train their model to predict the humans' trajectories. They let the LSTM network join and share information with LSTM networks in its range. Shi et al. [13] suggested LSTM networks that use encoding and decoding, which in turn, encode movements and interactions for a long sequence of predictions. Gupta et al.[4] suggested Generative Adversarial Networks that use encoding and decoding structure to predict future paths and avoid the existence of more than one prediction. Manh et al. [8] shows two models of Scene-LSTM that can predict human motion; it presented how the information from the scene is important by feeding it to the cells, which in turn, uses only the useful data to forecast next movements. Xue et al. [14] uses the Bidirectional LSTM to class people's destinations, which in turn, improves prediction precision.

Necessary crowd safety, such as at religious gatherings, concerts, or sporting events, can be improved by analyzing crowds behavior and improving the designs of crowd movement at large events [7, 6]. Crowd behavior and movement have been defined as two types: structured crowds, where people are heading in specific directions; and unstructured crowds, where people's directions cross each other[12]. Yamin proposed an app for crowd management [16] at the Hajj (an Islamic ritual), where most of the crowd formed as a structured crowd. In Switzerland, there is a festival that takes place at Z¨uri F¨ascht in Zurich; Ulf Blanke et al.[3] deployed an app for crowd management with sign points for each booth that must to be collected by the visitors. The RFID-based Hajj management system proposed by Al-Hashedi et al. [5] to manage crowds used data sharing. In addition, [9] managed the crowd by proposing using RFID with network communications. Since the Hajj is the largest religious gathering of people and behaves as a structured crowds, Yamin et al. [16] proposed integrating social media and mobile apps for Hajj management. A framework has been suggested for Hajj management by Yamin [15] to improve crowd motion research. He proposed a framework for monitoring hajjies(people who practice the Hajj ritual) upon their arrival at the airport, which is the starting point of their ritual participation. Additionally, Nidal Nasser et al [10] proposed a crowd monitoring and management framework for the Hajj gathering.

## 3. Methodology

Dense crowds usually form particular patterns depending upon the crowd type. According to [12], one direction and one goal is the pattern formed by a structured crowded, and different directions and different goals is the pattern of an unstructured crowded. Figure 1(b) and Figure 1(a) show examples of how structured and unstructured crowds. In a structured crowd, people usually maintain their direction in a particular pattern but adapt the path they use. By contrast, in an unstructured crowd, people create more than one pattern, but usually maintain their direction in a specific manner. The idea of an individual's "cone of vision" was applied in [2], and emphasizes determining the three closest people (near, mid, and far), then letting machine learning learn how these closest people in the cone of vision might impact the decision of an individual's choice of path. Figure 3 shows how [2] uses the "Unit Angle," which numbers the neighbors in the cone of vision. In contrast, LSTM has been used to predict people's trajectories based on past positions, such as in [1].

### 3.1. Overview on the LSTM

To enhance the performance of RNN to alleviate the difficulty of learning a long data sequence, LSTM has been suggested to fix the problem of vanishing and exploding gradients. Regarding vanishing means, the gradient tends to be smaller and smaller when we return to the earlier layers. By contrast, when we consider the exploding means, the gradient tends to get larger and larger when we go back to the earlier layers. In general, that may mean there is no problem with a small number of hidden layers, but it may cause an unstable situation when we deal with a large number of hidden layers. Some of the LSTM advantages are remembering data for a long time, and predicting more precisely sequential information based on previous data. By using three gates (forget gate, input gate, and output gate), the idea of an LSTM network can be formulated. Figure 2

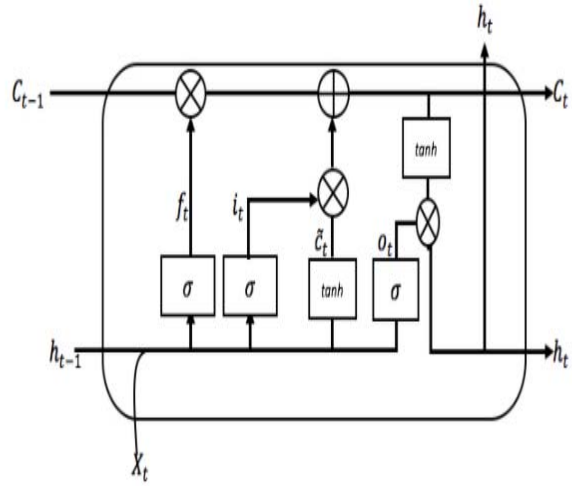


Figure 2. LSTM network example, which $f$ represents the forget gate, $i$ represents the input gate, and $o$ represents the ouput gate

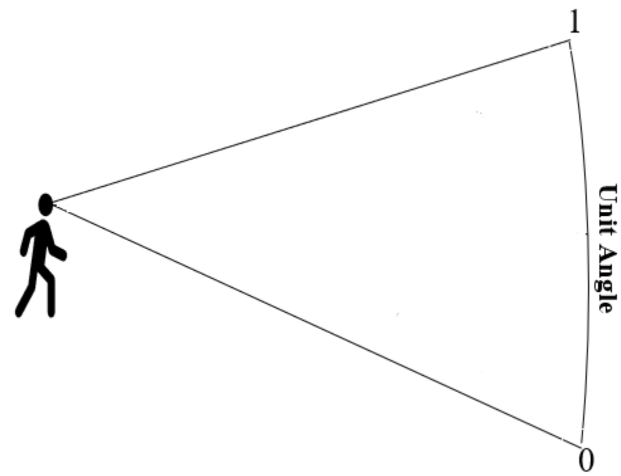


Figure 3. our model focuses on the individual's previous direction in his/her cone of vision to predict his/her next positions. The patterns of his/her last directions are represented as a distance and angle. The angle will be scaled between 0 and 1.

illustrates how an LSTM network functions in the LSTM modules.

### 3.2. problem statement

Our goal is the accurate prediction of the people locations from a sequence of data based on people's walking behaviors. The challenges always depend on the accuracy of the prediction for pedestrian trajectories in compare between the actual location with the predicted location. Our proposed model is based on two main factors: 1) the direction that the individual decides for his/her path results from the sequence of previous steps; and 2) the velocity which can estimate the final predicted location over time is based on the average of the previous speed. In our proposed model, we picture the behavior as a drawn pattern. By taking the notion of the field of view, we will draw the sequence of steps based on the Unit Angle, shown in Figure 3, for each individual. In other words, every person will have a sequence of data, each datum represents a value between 0 and 1 that denotes the direction of his/her cone of vision, which in turn produces a value for each step in the sequence. The velocity for every individual is an important matter in our prediction model in order to calculate distances more accurately. The prediction for speed in [2] was based on the last-step velocity. In contrast, our proposed model uses the average speed for every person that is calculated from a sequence of data. To process all these data, we will use LSTM networks to handle the sequence of data for pedestrians. LSTM networks have the proven ability for predicting sequential data. We employ the concept of the Unit angle discussed above, to get the directions of past trajectories, and then we feed the sequential results to the LSTM networks. The angle of direction in the cone of vision will be calculated as $\theta = Atan2(y2-y1, x2-x1)$, and the equation for distance is $Distance = sqrt((x2-x1)*(x2-x1)+(y2-y1)*(y2-y1))$.

### 3.3. Datasets

We extended our work in [2], which has two categories of datasets. The datasets have been acquired from the Netlogo simulation machine. One of the dataset represents structured crowded areas, in which people walk in the same way and go toward the same destination. The other dataset represents unstructured crowded areas, in which people go to several destinations in more than one way. Both of the datasets applied something similar to the social force model to simulate the behavior of people in a dense crowd. These social force rules include people keeping a fair distance between themselves and other people, between themselves and obstacles, acceleration of pedestrians to the desired speed, and the effect of the surrounding attractions. The datasets in [2] deal with the closest people in the cone of vision for each step that is fed into the neural networks. In short, the closest people's positions will be the inputs for the neural networks, which, in turn, allows for the prediction of the next direction. By contrast, using the same datasets, we will use past directions for every individual to obtain data about the patterns of his/her movements that , will predict next trajectories.

## 4. Conclusion

In this paper, we present a novel model based on an LSTM network for the prediction of pedestrian trajectories. Our proposed model depends on two factors, the last directions' values for each individuals, and the average speed for each individual. In other words, the model learns from

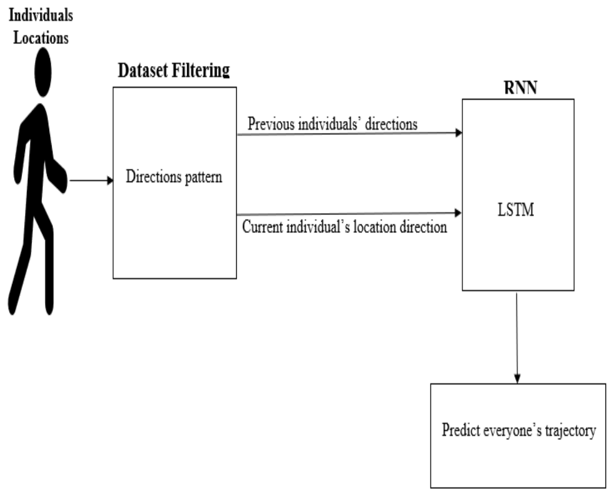


Figure 4. The workflow presents how to generate the Individuals locations data to be fed for the LSTM networks to predict the next individuals locations.

last values of directions and predicts the future trajectories by LSTM networks. Additionally, the model calculates the average speed for each individual from the last steps' velocity, then specifies the future locations of individuals. The experiment in the proposed model was applied to two kinds of datasets based on crowd type: structured crowded areas and unstructured crowded areas.

## References


[1] Alexandre Alahi et al. "Social lstm: Human trajectory prediction in crowded spaces". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 961–971.

[2] Alajlan, Edris Heckndorn, and Soule. "Using Neural Networks and Genetic Algorithms for Predicting Human Movement in Crowds". In: *The 22nd Int'l Conf on Artificial Intelligence*. Springer. 2020.

[3] Ulf Blanke et al. "Capturing crowd dynamics at large scale events using participatory gps-localization". In: *2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*. IEEE. 2014, pp. 1–7.

[4] Agrim Gupta et al. "Social gan: Socially acceptable trajectories with generative adversarial networks". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 2255–2264.

[5] Abdullah Hussein Al-Hashedi et al. "RFID applications in Hajj management system". In: *2013 IEEE International Conference on RFID-Technologies and Applications (RFID-TA)*. IEEE. 2013, pp. 1–6.

[6] Anders Johansson et al. "From crowd dynamics to crowd safety: a video-based analysis". In: *Advances in Complex Systems* 11.04 (2008), pp. 497–527.

[7] Barbara Krausz and Christian Bauckhage. "Loveparade 2010: Automatic video analysis of a crowd disaster". In: *Computer Vision and Image Understanding* 116.3 (2012), pp. 307–319.

[8] Huynh Manh and Gita Alaghband. "Scene-lstm: A model for human trajectory prediction". In: *arXiv preprint arXiv:1808.04018* (2018).

[9] Yamin Mohammad and Yasser Ades. "Crowd Management with RFID & Wireless Technologies". In: *Proceedings of First International Conference on Networks & Communications, IEEE Computer Society Washington, DC, USA*. 2009.

[10] Nidal Nasser et al. "An expert crowd monitoring and management framework for Hajj". In: *2017 International Conference on Wireless Networks and Mobile Communications (WINCOM)*. IEEE. 2017, pp. 1–8.

[11] Ovgu Ozturk, Toshihiko Yamasaki, and Kiyoharu Aizawa. "Detecting Dominant Motion Flows in Unstructured/Structured Crowd Scenes". In: Aug. 2010, pp. 3533–3536. DOI: 10.1109/ICPR.2010.862.

[12] Mikel Rodriguez, Saad Ali, and Takeo Kanade. "Tracking in unstructured crowded scenes". In: *2009 IEEE 12th International Conference on Computer Vision*. IEEE. 2009, pp. 1389–1396.

[13] Xiaodan Shi et al. "Pedestrian trajectory prediction in extremely crowded scenarios". In: *Sensors* 19.5 (2019), p. 1223.

[14] Hao Xue, Du Q Huynh, and Mark Reynolds. "Bi-prediction: pedestrian trajectory prediction based on bidirectional LSTM classification". In: *2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. IEEE. 2017, pp. 1–8.

[15] Mohammad Yamin. "A framework for improved Hajj management and future research". In: *ENTIC Bull* 2.08 (2008).

[16] Mohammad Yamin, Hasan M Al-Ahmadi, and Abdullah Al Muhammad. "Integrating social media and mobile apps into Hajj management". In: *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*. IEEE. 2016, pp. 1368–1372.