Journal of Scientific & Industrial Research
Vol. 82, October 2023, pp. 1009-1021
DOI: 10.56042/jsir.v82i10.3571

Navigation with a Cooperative Social Robot for a Group of Visitors using Face Detection and a 'Stop and Wait' Scheme

Vaibhav Malviya[1]* & Rahul Kala[2]

[1]National Institute of Technology Mizoram, Mizoram 796 012, India

[2]Atal Bihari Vajpayee Indian Institute of Information Technology and Management Gwalior, Madhya Pradesh 474 015, India

Received 12 February 2021; revised 12 June 2023; accepted 13 September 2023

The work is on the use of a robot as a guide that takes visitors for a guided tour around a facility. A past project of the research group proposed a robot guide that enacted a pre-recorded tour, however, had a limited applicability as the humans rarely followed the guided tour and the robot did not respond to the movement of the visitors. A robotic guide must ensure that it takes the visitors along, who are not left behind, while typically maintaining comforting distances from the visitor. The width and height of the human face is computed which is inversely proportional to the distance of the person from the robot. Further, we construct a method which guides visitors cooperatively. The robot moves sequentially to different locations with the visitors, and if any visitor is found missing, the robot stops and waits for that visitor. When the visitor becomes visible, the robot resumes the journey. The robot moves and navigates as a guide for a group of visitors maintaining appropriate distances from the visitor using the distance measurement methodology. The results are demonstrated by making the robot take visitors on a guided tour of the Robotics and Machine Intelligence Laboratory. The robot waits if a visitor leaves the group for calls or any other reason, while also waits if the visitors lag behind. The work demonstrates the ability of a robot to be socially complaint while taking a group of visitors on a guided tour.

Keywords: Behavioural robotics, Convolution neural network, Group detection, Robotic tour, Service robot

Introduction

Defining the behaviour of the robot is a very important and challenging, when operating with humans. Our aim is to make a robotic guide to take visitors on a guided tour at several places like tourist places, shopping malls, roaming at interesting places etc. The application area of our work is tourist places where visitors are interested for roaming and very excited to know important and historical facts about wonderful places. Currently, guides operate in these areas, who know the different sites of interest and take visitors on a guided tour. A robot should be able to render the same services socially. The robot needs to operate cooperatively to guide a group of visitors from the starting point to several visit sites in a sequential manner. Here the robot is a leader, and works as a multi-visitor tour guide. In the developed algorithm, the robot detects the visitors and moves from one place to another place. If any visitor is missing, the robot stops and waits for the visitor; and resumes when all missing visitors are located. The robot guides a group of visitors from a specified starting point to several labelled sites of visit. To perform this task, it is important to see the visitors, which is facilitated by a rear looking camera. The robot moves as long as all visitors are detected. When the robot reaches a site of interest, it stops and explains all important facts about that spot, followed by this the robot again proceeds for its next visit site and explains everything about that spot. The robot visits all sites of interest sequentially, waiting for the missing visitors, if needed.

The other major challenge of the work is to make the robot maintain appropriate distances with the visitors. This means that if the visitors are lagging behind, the robot must wait before continuing the tour. First a set of socialistic experiments were performed aimed to find the distance that the human subjects find socially appropriate with the robot. Using a number of test subjects, there is no single distance that can be considered as a gold standard, however, the majority of the comfortable distances in such settings came to be in the range 2.1–3.5 meters. The distance to the human can be measured by proximity sensing, which will be accurate but incurring uncertainties whether the distance reported is of the human face. Alternatively, stereovision may

*Author for Correspondence
E-mail: vaibsidea@gmail.com

be used to get the distance, incurring stereo correspondence uncertainties. Assuming the visitors to be of similar heights and with similar face dimensions, the problem can be much easily solved by a single monocular camera alone. The dimensions of the face are inversely proportional to the distance of the subject from the robot. So, the robot only detects visitors in the range 2.1–3.5 meters. To do the distance measurement, we found width and height of every face at a multiple distance values and performed the experiment with multiple test subjects. The approach uses a Convolution Neural Network for the detection of the face. The face detection scheme is able to detect a face at any scale and also reports the width and height of the detected face.

The laboratory of the authors attracts several visitors throughout the year, due to which a past project of the research group envisioned a robot taking the visitors on a guided tour of the laboratory. The robot designed had a pre-recorded tour that the robot could enact on every visit. However, practical demonstrations suggested that the visitors rarely followed the robot as desired, while the robot was not responsive to the attention of the visitors. The social robot navigation requires information that the visitors are exactly following and available behind the robot or not in the 3D real-world environment. When the human is following the robot then it is not necessary that every time the human can place himself/herself exactly behind the robot. If they are not present behind the robot, in such cases the robot must stop and wait for the visitor to appear behind it. As the visitors appear and are detected by the robot, the navigation module should resume its journey. A human tour guide in such situations reacts to people. In other case, suppose some people are still at the previous site or cannot catch up, the human tour guides stop, and the robot should have the same social etiquettes.

Here ‘stop and wait’ scheme is developed for the execution of our navigation module. Our stop scheme calculates the distance between robot and humans based on the bounding box which is plotted on the face, and it is inversely proportional to the distance. If the height and width of the rectangular box is too small, it indicates that the distance between visitors and robot are too much and thus the visitors are not following the robot and still present in somewhere in the 3D real-world environment. Therefore, the robot will stop and wait for the visitor to catch up. Moreover, as the height and width of rectangular box is increased, distance between visitors and robot also decreases, resultantly, the visitor has appeared behind the robot and is following the robot and the robot continues its journey.

The algorithm is tested on the Robotics and Machine Intelligence laboratory of the institute. The Pioneer LX robot is used for the experimentation. The visit sites for the robot are the same as used by the lab in-charge for taking the visitors allowed. A few sites are added in addition. The algorithm is performed on different group sizes. The laboratory accounts for different motion planning challenges including factors that including different structure of the environment (like open area, corridors, inter-sections). The visitors are asked to have different inattention as the robot operates, requiring them to stay away from the desired path. The robot is first localized to the initial point of the journey. Visitors are then asked to stand at the rear side of the robot. If initially all the faces are detected then the robot starts its journey, otherwise it waits for all the visitors that should be a part of the tour. The experiments are done on the Robot Operating Systems Library. The library allows for the face detection, motion planning and the scheduling of visit sites modules to run in parallel and operate via communication with each other so as to have an overall seamless working of the algorithm.

We propose a strategy for a social robot tour guide navigation scheme in a real-time environment. The main contributions are:

(i) Demonstration of robot as a human-aware guide for multiple humans using a budget limited field of view monocular camera to reduce high expense.

(ii) Heuristic detection of a person leaving the group based on the current context and intends.

(iii) The socialistic human behaviours have been studied such as a person leaving the group, re-joining the group, following another person (when any human leaves the group and sees something at the previous visiting sites, then a friend will also leave the group and go towards the friend).

(iv) We propose a new ‘stop and wait’ scheme, when any visitor leaves the group then the robot stops and waits for visitors to join in the group. If the person re-joins, the robot is able to detect them and continues its journey.

(v) The proposed approach accurately detects people leaving the group and performs a better socialistic navigation in the indoor robot service and application, unlike many current applications where humans are expected to religiously follow the robot and the robot does not react to people's movements.

Related Works

Recent research into human robot interaction in the area of cooperative robots has entered into service robotics such as a robot to serve a food in hotels, hospitals, and play games in a ground or picnic spot. Most of the recent research studies that a robot can participate in social human interaction as co-workers.[1–4] Different kinds of inspections have been performed for the human's attitude towards the robots and their awareness of robots, as an example robotic dog AIBO[5] an autonomous robot built for entertainment and attraction purpose. It has an ability to analyse annoying emotions, learning and developing its skills, can show its feelings, and can communicate and interact with human and children. Kerstin *et al*.[6] examined the people's perception and their ambition for cooperative and working robots. For social robots, the working environment can be dynamic and uncertain. The major intention of motion planner is to take into account the movements of the human in order to make safe and familiar paths to be followed. Pransky *et al*.[7] proposed different applications of a robot for the future by explaining the pros and cons of this type of working robot. The 'Robotic Bulter/Maid' accomplishes domestic task, but also causes problems at home making people feel inessential .A 'Robot Nanny' was also introduced that could play with the kids and grub them; however, also inhibited them from meaningful communication with human and made communication as ordinary. Vaughan *et al*.[8] designed a complete robot system with the capability to gather a flock of ducks and carry out stunt to safely deliver them to a predefined goal. Moreover, object detection algorithm [9] can also be utilized here for robotic application.

Various studies have been performed on the robots to guide people in hospitals,[10] railway-stations,[11] museums,[12] as assistants[6,13] and in shopping malls.[14] Sisbot *et al*.[15] examined and analysed robots planning routes as per human's preference. Suppose there are multiple ways from one place to another and people opt any specific route according to their preference, then the robots are also able to understand the behaviour and account for user preference in their route. The robots do not only find the safe path but also plan socially acceptable routes for the people. Robots working around the humans should move in a simple and clear way, so that they do not endanger the people in their company. Shiomi *et al*.[16] developed a group attention control system for the robots to interact with the people in a group.[17] If the robot moves in a backward or forward direction as the people move around it, then the human robot interaction also changes.[16] In a recent paper,[18] a hybrid framework has been designed to avoid humans according to social conventions which were followed by humans, keeping a socialistic distance when avoiding humans.

Martinez-Garcia *et al*.[19] performed a qualitative analysis of the motion of people and multiple robots. But this system is not applicable in a realistic situation such as presence of obstacle, or straying away from their group. These challenges and motion of the autonomous robot are handled by various heuristics that creates a realistic system. Garrell *et al*.[20] built the robotic workplace, in which the robot had to operate a guidance task in an open and unbounded area in the presence of obstacles. The algorithm could also find the orientation, velocity, and position of the people and robots, as well as position and availability of obstacle in that workplace. It could also predict the intersection of the human with obstacle and detect the person who is leaving or straying the group using a Particle filter.[21] It was also capable of finding the trajectory of the robot to attain the goal while preventing people from leaving the group. This work focuses on face detection and its application in the navigation of the robot. In general, face detection has been limited to the frontal face only with respect to the face features like eye, lips, cheek, hair etc. Beyond this some authors have proposed a methodology of face detection that includes profile faces.[22–25] Two methodologies have been followed in the previous research for pose. The first methodology is based on the representation of faces individually and comparing two faces as a set of images. The second one is to represent the whole track, like 3D model, or a single vector or other manifold, with no accurate reference to separate face detection. A very good literature review of these approaches is given in face recognition's review paper.[26] The first methodology is most followed and the representation allows to develop directly on the face image. Some standard methods compute the feature around the facial landmark and add to a form of descriptor vector for every frame.

These are classified by the min-min distance on the set of descriptors. Among this, several methodologies have also been proposed in the literature, some methods use HOG (Histogram of Oriented Gradient), SIFT (Scale Invariant Feature Transform), spatio-temporal gait measurement system.[23,25,27–30] Activity of the people is also a major concern in an extensive variety of application such as human computer interface, video surveillance, face detection, management of face image database[31] and so on. Detection of faces is the decisive step in these identification applications. Further face detection and tracking analysis with different height of the camera has been carried out in the paper,[32] here camera is mounted on the robot and apart from this optimal balancing[33] of the robot is also important for pre-defined path. Further robot can also perform behaviour-based tracking of humans using a particle filter for a better decision making.[34,35] To make the mobile robot social, real human walking characteristics are considered such as distance variation, velocity etc.[36] Similarly, learning based approach is helpful for replicating the human walking behaviour by preparing a rea-human trajectory database using a 3D lidar.[37] Behaviour analysis in traffic is similar to indoor environment and helpful for proposing a risk modelling for risky overtaking behaviour that affects traffic safety.[38]

Research Gap

From the literature, it is evident that the problem of socialistically guiding a group of visitors has not been studied from a navigation perspective. The current studies primarily study the interaction between the robot and human only. The behaviour of the visitors affects the behaviour of the robotic guide, which is important to taper.

Materials and Methods

Overall Solution Design

The work introduces a novel strategy for the navigation of visitors with a cooperative robot. Inattention is a common problem with modern tours wherein visitors temporarily leave the tour guide for different reasons, prominently to attend mobile calls, they get attracted to other places, attending kids, here, we have considered such circumstances in which visitors move away from the group and join the group at a later stage. The aim is to make the robot wait at such circumstances.

An interesting case comes when the robot, travelling as per the optimal path from one site to another, encounters a corner to navigate wherein the robot makes a sharp turn. If the robot takes the turn, then the camera which is mounted on rear of the robot also makes the same turn, while the visitors following the robot will account for the turn a little later. Due to this the robot might not see the visitors and anticipate that visitors are missing. To tackle the problem a time delay is added and the robot only makes the stop when the visitors are not visited for a threshold of time. This also accounts for temporary inattention, noise, and similar factors. However, it is important that if the robot is moving and taking turns, while planning its path from current position to the desired visit site, then humans should also take a turn following the robot without much lag, otherwise the robot will wait for the visitors.

To make the robot move, we need to know the map of the workplace, so that the robot can continuously plan the optimal path from the current position to the next visit site and navigate the visitors in a group. The complete tour can commence by visiting all sites in a sequential order. The face detection happens by the use of Convolution Neural Network. Initially the robot is at the start place and waits for all visitors. The tour starts when all visitors are found at the rear of the robot. If the robot successfully reaches the desired visit site, then it stops and interacts with the humans in a group and explains every important fact about the visit site. After telling important facts of that place, the robot will again start the journey and navigate the visitors to the next visit site. In these scenarios one interesting aspect of the algorithm is that suppose once the robot has found an optimal and safe path form one place to another visit site, and it has started the tour. Consider that after some time an obstacle appears on the path, then the robot will again re-plan the path. However, this path may be different from the previous path. This behaviour is also displayed by human guides who change the path to a visit site depending upon the congestion levels. Another aspect of re-planning is applicable when the robot has found a safe path from one place to a visit site, but during run there happens to be a new obstacle making the site unreachable. As an example, a new obstacle place leaves no space for the movement of the robot as well as visitors, then the robot will move only till a safe distance of the visit site and after reaching this distance, the robot will again stop and visualize the situation of the environment. If the visit place is unreachable safely, then the robot will not move forward and will stop its journey. So, the robot is aware of the safety in the environment or workplace. The basic working methodology of the approach is shown in Fig. 1.

Robot Navigation

The other module of the work is the navigation of the robot as per the principles laid out. A requirement is that the robot maintains a socialistic distance from the visitors. In case the same is not the case, the robot must wait as the visitors may be lagging behind. The waiting behaviour is the same irrespective of whether the visitors are not visible or when the robot is not maintaining a socialistic distance. Let d_{soc} be the distance that must be maintained between the robot and the visitors. A realization from the socialistic experiments is that this distance varies by a very large factor between different visitors, between different moods of the visitors, and between different visit sites. Further, not all visitors may be at comforting distances and will have mutual socialistic forces between each other, which affects the placement and hence the distance. Since the socialistic distance d_{soc} can only be approximated within a range $d_{soc} \sim [d^{min}_{soc} d^{max}_{soc}]$, it is suggestive to have an estimate of the distance between the robot and the visitor. The closer is the face to the camera (and the robot), the larger is the area of the face and vice versa. So, the area of the face is inversely proportional to the distance of the visitor from the robot. Further assuming the aspect ratio of the human face to be nearly constant, the length and the width of the face is inversely proportional to the distance. This means that = the limits of the socialistic distance $[d^{min}_{soc} d^{max}_{soc}]$ can be translated into limits on the length of the face $[\eta^{min}_{x}, \eta^{max}_{x}]$ in the image and the width of the human face $[\eta^{min}_{y}, \eta^{max}_{y}]$ in the image. The best mechanism to understand the navigation principle is to model the behaviour as a Behavioural Finite State Machine. The modelling enables model different behaviours that the robot displays at the different phases of time, and to further model the conditions over which the robot behaviour shows a transition. The complete model is shown in Fig. 2.

Algorithm 1: Face detection Algorithm

1. Initialize P-Net, R-net, and O-Net
2. Import and initialize the camera
3. Do
a. I ← Image from camera

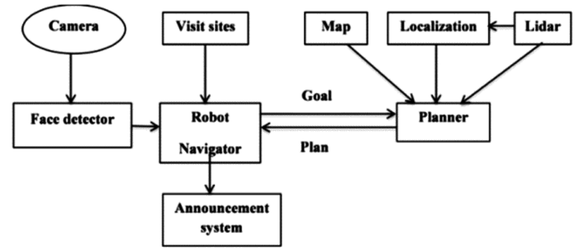

Fig. 1 — Overall Architecture of the system

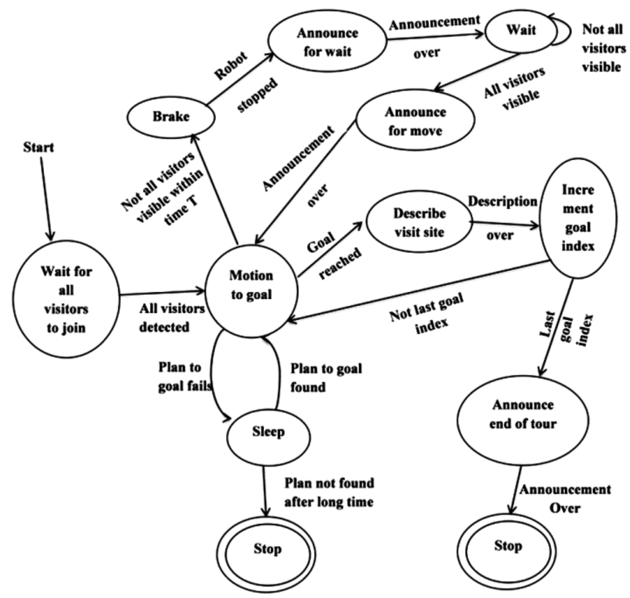

Fig. 2 — Behavioural Finite State Machine Representation of the system

b. F ← Detect the face by the convolution neural network which is P-Net, R-Net, and O-Net on I with minimium size $[\eta^{min}_{x}, \eta^{min}_{y}]$
c. F ← Select faces strictly within the image
d. if (length(F) = noVisitors∧ (width(f) ≥ η^{max}_{x}∧ height(f) ≥ η^{max}_{y}) ∀ f ∈ F
i. publish 'all visitors visible'
e. else
i. publish 'all visitors not visible'
f. Display image I and faces F
4. while (state ≠ stop)

Algorithm 2: Robot motion algorithm

1. sites ← sequence of sites to visit
2. siteIndex ← 1
3. state ← 'wait for all visitors to join'
4. while 'all visitors visible' not received, sleep for a brief time
5. publish site[siteIndex].location as Goal
6. while (state ≠ stop), sleep for a brief period

Algorithm 3: Goal Reached Behaviour(result)

(callback called immediately when a goal is reached or plan fails, result stores the planner's output)
1. if result = "success"

a. state ← ‘describe visit site’
b. Announce site[siteIndex].announcement
c. while announcement is not over, sleep for a brief period
d. state ← ‘Increment goal index’
e. siteIndex ← siteIndex + 1
f. if siteIndex≤length(site)
i. state ← ‘Motion to goal’
ii. publish site[siteIndex].location as Goal
g. else
i. state ← ‘Announce end of tour’
ii. announce end of tour message
iii. while announcement is not over, sleep for a brief period
iv. state ← ‘Stop’
h. Reset failTimer
2. else if result = “fail”
a. state ← ‘sleep’
b. update failTimer
c. if failTimer≥T_{fail}, state← ‘Stop’

The development methodology is to use Service Oriented Architecture due to the modularity of the different components and naturally suiting the nature of the Robot Operating System, the framework used for the development. The application of Service Oriented Architecture does not easily translate to the Behavioural Finite State Machine. Further, many linear components of the Behvioural Finite State Machine are combined into procedural units while developing. Many components use the face detection results as sensory information. Algorithm 1 gives the face detection algorithm.[39–41] Algorithm 2 implements the initial motion of the robot. The motion thereafter is by a set of registered call backs only. Algorithm 3 implements the behaviour when the goal is reached. The motion planning library, not a part of the algorithm, calls the callback function whenever the goal is reached successfully or the plan fails. The result stores the results received from the planner. The other module is to stop and wait when all visitors are not visible and to move otherwise. The face detection algorithm keeps publishing the visitor status and therefore the corresponding callback that receives the visitor status is always active, where the logic is implemented. This can also be visualized as a continuous loop that maintains the state and invokes the corresponding behaviour. The callback implementation is given by Algorithm 4.

Algorithm 4: Waiting Behaviour (faceDetection)

(callback called immediately and every time when face detection messages are received)

1. if faceDetection ← ‘not all visitors not visible’ ∧ state = ‘Motion to goal’ ∧notFaceVisibleTimer> T
a. state ← ‘Brake’
b. publish ‘cancel motion’
c. while(velocity>0) sleep for a brief period
d. state ← ‘Announce for wait’
e. announce wait message
f. while announcement is not over, sleep for a brief period
g. state ← ‘Wait’
2. if faceDetection ← ‘not all visitors not visible’ ∧ state = ‘Wait’
a. state ← ‘Announce for move’
b. announce move message
c. while announcement is not over, sleep for a brief period
d. publish site[siteIndex].location as Goal
e. state ← ‘Motion to goal’
3. if faceDetection ← ‘all visitors not visible’ Reset notFaceVisibleTimer
4. if faceDetection ← ‘not all visitors not visible’, update notFaceVisibleTimer

Implementation Details

All the methods and algorithm are implemented and tested on the Robot Operating System (ROS) framework. The algorithms are tested on a 16GB RAM, CORE I7 processor, and Linux operating system. The camera is required to visualize and capture a video in real time. So, setup of the camera for real time inputs and importing it in the program at execution time is also needed. Tensorflow is used to visualize and detect the human faces. Since the robot has to move so localization and load an environment map is very important task. MobileEyes is used to initially localize the robot and load the environment map. Through the localization, robot is able to know its current position in the environment. MobileSim is a simulator used for developing the algorithm on the simulator, which is finally tested on the real robot. ROS ARNL library provides the basic functionalities of operating the robot.

Pioneer LX is used for the experiments (Fig. 3), which is based on the autonomous indoor vehicle (Adept Lynx AIV). Modular programming can be done on this robot since it is programmable. This robot has a capability to know its current location and can plan a path from one place to another. The robot can move with maximum speed 1.8 metre/second and this robot also contains 2 drive wheels, 4 coasters and the swing radius is 13.5 inch. To visualize the human and to detect the face webcam is used as a camera.

Fig. 3 — Pioneer LX robot and its setup for navigation

Webcam is connected to the Pioneer Lx robot and it is fixed at a top of the pioneer robot so that it can move along with robot. All the images and video of workspace and humans are detected and visualized by this webcam which is fixed and mounted at the top of the Pioneer LX robot. The robot analyses the humans and their faces by this webcam while the robot is moving at the appearance of humans as well as becomes static at the nonappearance of humans. Here the resolution of the 2D image plane of the camera is 640×480. The robot localizes using the position encoders and a LiDAR sensor. The robot uses the optical quadrature shaft encoder. The encoders have a resolution of 9,550 ticks per wheel revolution that translates to approximately 30 ticks per millimeter. The LiDAR sensor used for localization and navigation has a resolution of 0.5 degrees.

In this study we are working for navigation of visitors with the robot. We have used the Robotics and Machine Intelligence Laboratory located at IIIT Allahabad as an environment for the robotic tour. The robot needs to navigate with the visitor to visit different sites in the environment. So it important to know the exact positions of all visit sites which would be visited and navigated for the robot. We create a map for the workplace where the robot will be move and navigate the visitors. Map is created by using the Pioneer LX robot. We manually traversed the robot by joystick in all the area where robot will automatically move and navigate the visitors. The environment is scanned by manually traversing the robot and at the end of the traverse the lidar log file is generated which is processed by Mapper3 library to convert to a map as used by the robot. The map is shown in Fig. 4. In Fig. 4 all the boundaries are covered and can be seen as black coloured boundary, a place where the robot cannot enter and obstacle are also sketched as black colour. Here robot is represented as a red coloured ball and the path traversed by the robot during map creation is drawn as blue colour trajectory in the Fig. 4. Fig. 5 represents the map file of the environment where the robot will be moving and navigating the visitors. In Fig. 5, all visit sites are annotated. The robot starts its journey from the starting point. Robot detects the visitors from this starting point and navigates the visitors sequentially on all visit sites.

Results

Socialistic Experiments

An aim of the work was to make the robot maintain the same distances that a guide would typically maintain with the visitors. However, the robot forms a reasonably different socialistic class and therefore it is not necessary that the visitors feel comfortable at the same distance with the robot as they would do with the human guide. Therefore, the first aim of the experiments was to find out the distance that the robot should make with the visitors. The problem with the specific work is that the robot is guiding a group and not a single visitor. Therefore, even if a visitor finds some distance comforting with the robot, it is not necessary that the same distance shall be actually possible considering other members of the group. The second aim of the socialistic experiments is to find out how do the socialistic distances change for small to a fairly large group size.

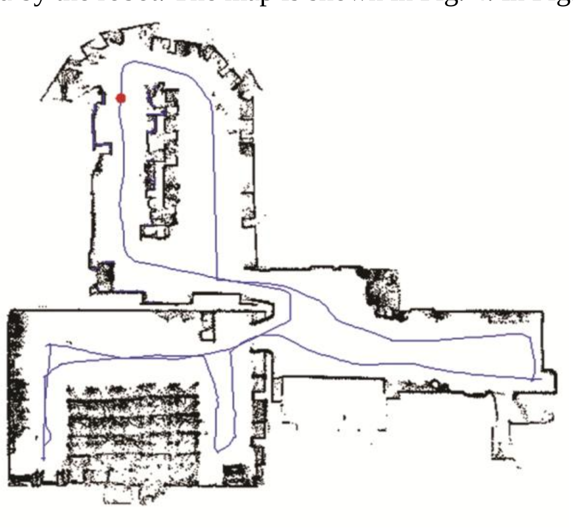

Fig. 4 — Initial map created by Pioneer LX robot

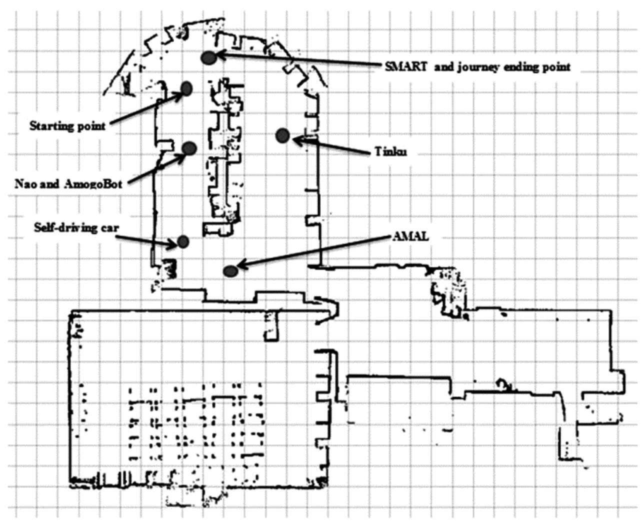

Fig. 5 — Map of robotics and machine intelligence laboratory indicating the visit sites

In the initial set of socialistic experiments, the visitors were asked to stand at a comfortable distance from the robot. The visitors showed a great disparity in maintaining distances, ranging from fairly small to fairly large. This makes it clear that there is no single socialistic distance and different visitors have different socialistic distances. A part of the problem was also because of the fact that since the visitors had never travelled with a robot before, they did not have any preferences which in case of humans are developed after a very long time. Therefore, some visitors got very connected and stood close, while the others were fearful and preferred to keep a distance. The last aim of the socialistic experiment was hence to get the range that appreciably covers the acceptable socialistic distances that the visitors keep with the robot and to convert the same into the width and height of the faces as seen in the camera attached to the robot. To explore the socialistic distances between the robot and visitors, while the robot is guiding in a group, we have calculated the height and width of all the faces in different scenarios and multiple distances with multiple visitors. We determined the height and width of only one visitor when the visitor is at multiple distances from the robot like 1, 1.5, 1.8, 2, 2.5, 2.7, 3.1, 3.4, 3.5 meters and so on. Similarly, height and width of all face with one, two, three, four, five, and six visitors in a group have been determined and tested with multiple distances from the robot. Suppose there are five visitors in a group, and distance between the visitors and the robot are 1.1 meters, 2.1 meters, 1.6 meters, 3.4 meters and 2.8 meters. In this case we calculate the height and width of all these five faces the distances from the robot. In each experiment the visitors are asked whether, given the group and the position of other members of the group, they find the distances comfortable or not.

We applied this strategy with different number of visitors in a group (like 1, 2, 3, 4, 5, 6) with various distances (like 0.92, 1, 1.1, 1.4, 1.8, 2, 2.4, 2.7, 3.1 meters etc.) to determine and analyse the height and width of all faces. By doing this we find a lower and an upper bound of height and width of the faces which would be applicable to the robot to navigate the visitors in a group. The best estimate of height and width when the robot's decision to move matches with the comfort as reported by all the visitors is considered.

All the experiments have been summarized in Table 1 and supplementary table. In these tables there are several scenarios with multiple test cases. Here every tables contain 9 column, where first column represents a unique scenario ID for every set of different experiment with different number of visitors with different distance from the robot, second column is the image captured from a rear camera of the robot, total number of visitors in a frame is actually present have shown in the third column, number of visitors detected by the robot is shown in the fourth column, fifth column represents the visitor ID for every visitor

Table 1 — Distance Measurements Group 1: Computation of width and height of visitors at different distance from the robot

Scenario ID	Number of visitors (actual)	Number of visitors (predicted)	Visitor ID	Weight of face	Height of face	Distance from robot	Comfort factor
1	2	2	1	15	19	3.18	1
			2	17	21	2.95	
2	2	2	1	17	21	2.95	1
			2	15	18	3.36	
3	2	2	1	16	19	2.95	1
			2	24	30	2.34	
4	2	2	1	16	20	2.95	1
			2	31	37	1.81	
5	2	1	1	17	21	2.88	0
6	1	1	1	11	14	4.21	0
7	2	2	1	19	23	2.95	0
			2	12	14	4.21	
8	2	2	1	17	21	2.88	1
			2	15	19	3.58	
9	2	2	1	27	32	2.10	1
			2	16	21	2.96	
10	2	2	1	37	48	1.37	0
			2	27	33	2.10	
11	1	1	1	14	16	3.97	0

in the group, sixth and seventh columns represents the width and height of human face corresponding their visitor ID, distance (in meter) between robot and visitor is represented in eighth column, and the last column represents whether the robot should move or comfort factor. Here we have used two symbols [0, 1] to state the comfort factor. If the value is 0 then the robot should stop and wait for visitors because a visitor may be missed or may be far away from the robot, and if it is 1 then the robot has detected every visitor in the group and is moving towards its next visit site with the members of the group. Consider Table 1; there are 11 scenarios that have been tested with distinct number of visitors with different distances from the robot. Here in scenario ID 1, there are two visitors that are present in a group and each visitor has a unique visitor ID as 1 and 2. The robot detects both visitors via their faces and distance between visitors 1 to robot is 3.18 meter and visitor 2 to robot is 2.95 meter. So, the width and height of all face will not be the same here since both visitors are at distinct distances from the robot. The width and height of the face of visitor 1 is 15 and 19 while visitor 2 is 17 and 21 respectively. Also, in this scenario robot's comfort factor is 1, it means that the robot will be moving and navigating to all the visitors in a group. This complies with the personal observation that they were happy for the robot to move as a guide. Scenario 2 is also similar to scenario 1 with the robot's comfort factor being 1 even if the robot has visitor 1 and visitor 2 at the distances of 2.95 meter and 3.36 meter respectively, with different width and height of face. Here width and height of visitor 1 is 17 and 21 while visitor 2 is 15 and 18 respectively as determined by the robot. Moreover, scenario 3 and 4 are also the same as scenario 1 with robot's comfort factor as 1 and a distinct value of width and height of faces in a group. Consider scenario 5; here the comfort factor is 0, because in this image there are two visitors in a group while the robot detected only one visitor with a distance of 2.88 meter distance and determined 17 as a width and 21 as a height of that visitor's face. The second visitor's face was hidden by the first visitor and hence his visitor's face was not visible by the robot. In scenario 6, there is only 1 visitor is in the group (group size is 1) while the robot does not move and the comfort factor is 0, even if robot is detects this visitor because the visitor is too far away from the robot (4.21 meter) and width and height of face is also very small number which is 11 and 14 respectively. The visitor reported an uncomforting distance and the need for the robot to wait. So, if any visitor is at a very long distance from the robot then the robot will not move and wait for the visitor to come near to the camera. As the visitor comes near to the camera, the distance between the visitor and the robot decreases, and the width and height of the face increases. In the scenario 7 there are two visitors in the group and both are detected by the robot. However, robot is at a comfort factor of 0 since visitor 2 is detected with a very large distance of 4.21 meter. So, to take the robot in status 1, visitor 2 will have to reduce his distance from the robot while the robot should wait. The visitor confirmed that the distance was too large for comfort.

In the scenario 8, the robot has a comfort factor of 1 as its distance from visitor 2 is 3.58 meter, because width and height of face is within bounds. Scenario 9 is the same as scenarios 1, 2, 3, and 4. In scenario 10, the group size is 2 and both visitors are detected by the robot, but the robot comfort factor is 0 because height of the face of visitor 2 is too large and the visitor has a small distance from the robot (1.37 meter). The visitor was staring too close and it was regarded as unnatural. So in this case visitor 2 should move in the backward direction to decrease the height of the face to make the robot comfortable. Lastly, in scenario 11 with group size 1, the robot also detected the visitor but the robot was at a comfort factor of 0 because this visitor is detected from very long distance from the robot 3.97 meter and width of face is 14, which is also not a comfortable distance as judged by the visitor. Similarly the cases with other tables are analysed. After looking at all the experiments with different visitors, different groups and different distances with the robot and we have found the 15 minimum and 45 maximum as very good bounds for width and height of the faces, so as to maintain comforting distances with the robot. The robot detects the face, and moves towards the visit sites with the group if the width and height of the faces is lies in between the range 15 to 45. This range value (15 to 45) of width and height produces a comforting distance between visitors and robots as a minimum of 1.2 meter and a maximum 3.5 meter. So, while robot is navigating in a group, the visitor will have to maintain a distance between robot to 1.2 meter to 3.5 meter. Otherwise the robot waits for the visitors. In very less and exceptional cases the value of width and height of the face is larger than the upper

Fig. 6 — The robot guided tour (a) Robot is at its initial state or starting point of the journey and detected the visitors through their faces through the webcam mounted on it. (b) Robot reaches its first visit site, (c) NaO and Amihgo Bot are situated at the first visit site of the robot. (d) Robot has arrived at its second visit site, an autonoumous car. (e) Robot reached at its third visit site, AMAL. (f) Suddenly robot stops because a visitor is not visible here and waits for the visitor to appear in the camera. (g) Robot sees the visitors and detects their faces and makes an announcement to move. (h) Robot moves towards its next visit site, Tinku a socialistic robot for kids. The robot explains interesting facts about Tinku. (i) Robot moves towards its last journey point (j) Robot arrived at its last visit place; SMART (k) The robot terminates its tour successfully completes its robotic journey.

bound, when the visitor has moved too close to the robot. The robot will not move, which does not improve the situation, but the probability of this situation to arise is very less and the visitors will have to understand this social cue and move backwards. This is unlike the case when the visitor is far behind and the robot waiting gives enough time for the visitor to come and join the group.

Results on a Real Tour

The results of the algorithm are given in Fig. 6. The results are also shown in the supplementary video.[42] In Fig. 6(a) the robot detected visitors and started its journey and this point was also the initial or starting point of the tour. After a few milliseconds, the robot moved towards its first goal (Fig. 6(b)), where the NaO robot and AmigoBot robots were situated. The robot reaching the first visit site is shown in Fig. 6(c). Now the robot stopped and explained important facts of these two robots. After explaining everything of Nao robot and AmigoBot, the robot again started its journey while simultaneously detected the visitors. Followed by a small-time gap, the robot reached its second visit site where the mini self-driving car was parked. The robot stopped here and introduced the autonomous car to the visitors as shown in Fig. 6(d). The robot thereafter reached its third visit site where Adaptive Modular Active Leg (AMAL) was present and introduced important facts to the visitors in the group and moved to the next goal, shown in Fig. 6(e). However, after visiting AMAL, a visitor was not found and therefore the robot stopped here and made an announcement "*looks like I am missing someone, let us wait*" and the robot waited for the missing visitor as shown in Fig. 6(f). As the visitors appeared in the camera mounted on the robot, the robot detected the visitors and made an announcement "*Okay! I have found you let us move*". The robot restarted its missed and remaining journey. The robot detecting the visitors and moving towards its fourth visit site and simultaneously visitors are also following the path navigated by the robot is shown in Fig. 6(g). In Fig. 6(h), the robot arrived at its fourth visit site where *Tinku* robot was positioned. The stopped here and told interesting facts about the robot to the visitors in the group. Now only one place was left in the journey. The robot was ready to departure for its last goal as shown in Fig. 6(i). The SMART robot was located at the last visit site in this robotic tour. In Fig. 6(j), the robot successfully reached its last visit site. The robot stays here and gives the brief explanation of the SMART robot to the visitors. Followed by all these steps, the robot successfully completed its robotic tour and navigated the visitors, shown in Fig. 6(k).

Discussion

The proposed scheme is for small group tours only. The face detection algorithm has a large scalability and will continue to have a reasonable response time and is not a bottleneck. However, in large groups, it is impossible for all the visitors to come inside the same image. In tours, visitors typically get close to the guide and in case the group is large, the visitors spread all around. The limited field of view camera can thus only see a sub-set of visitors. This is an actual problem with tour guides in large groups as well, wherein counting the visitors only takes place at major stops by actually roaming around the group before departing. One of the major limitations is that currently the decisions are made based on the immediate percept of the robot only. The work needs to be extended to perform tracking for humans because every time detection is not possible due to illumination changes and image distortions due to a motion blur. Currently, while the experiments were carried out under realistic settings with the complete robotic setup including humans, the humans were detected by their faces which take some time that increases the response time of the robot. The current decision making is restricted to the presence or absence of the person only and is also limited to using a single limited field of view camera. We do not exact the human position in the 3D indoor environment. The experiments need to be done using additional sensors to determine the human position in the 3D indoor environment. The major application was enabling the robot to take an informed decision based on the face detection. Under these limitations, additional sensor requirements and its applicability need to be explored for a comprehensive navigation module. Currently the human intent is limited to leaving or rejoining the group, for small-sized groups with no occlusion. In the future, a visitor can be observed for some time and a new algorithm can be used that will be able to learn the distinct types of patterns or behaviours that the visitors form between themselves. The model will be able to identify and analyse which patterns are good or predict the motion of humans based on the learned intents and behaviours. This information can be used for a more informed decision making of the robot. In this way the robot may act as an expert guide that adapts its behaviour based on the perceived behaviour of the other visitor. There are several other directions where the proposed approach can be taken forward in the future. First is to model multi-robot multi-human groups, involving a group of robots that take a group of visitors on a tour. The robots may collectively not be able to detect all the people, while will have to make a decision based on the non-occluded visitors alone. The group may pass through regions of high density of crowd where several non-group people will come in-between the group that the algorithm should be able to handle. The other direction of work is to allow interaction between the human and robot where the human can ask queries to the robotic guide, while the robotic guide can adjust the tour as per the perceived emotions of the humans. Further, low-cost robotic guides will require a low-cost robot that can navigate based on low-cost localization system relying on local odometry and global markers in tourist places, alongside a low-cost vision system. Work for the development of such robotic systems needs to be looked into.

Conclusions

The study demonstrated a social robot that can take a group of visitors on a tour while stopping if the visitors went missing. A real tour was exhibited using the Pioneer LX robot at the Centre of Intelligent Robotics at IIIT Allahabad with visitors who did not religiously follow the robot. The study makes significant advancements toward the adoption of robots as tour guides while being socially responsive to visitors. The study demonstrates one of the several application-specific characteristics that a social robot in the service sector should have the remaining need to be addressed in the future. The algorithm needs to be extended to perform tracking to eliminate false negatives while perceiving the humans as per their 3D poses to make more informed decisions. Future work will incorporate large groups that can only be partially observed, while the robot operates in dense scenarios involving non-group visitors.

Acknowledgement

This work is supported by the Science and Engineering Research Board, Department of Science and Technology, Government of India vide research grant ECR/2015/000406.

Conflict of interest

The authors have no conflict of interest to declare.

References

1 Ishiguro H, Ono T, Imai M, Maeda T, Kanda T & Nakatsu R, Robovie: an interactive humanoid robot, *Ind Rob*, **28(6)** (2001) 498–504, doi: 10.1108/01439910110410051.

2 Hornby G S, Takamura S, Yokono J, Hanagata O, Yamamoto T & Fujita M, Evolving robust gaits with AIBO, *Proc Int Conf Rob Autom* (IEEE) 2000, 3040–3045, 10.1109/ROBOT.2000.846489.

3 Nourbakhsh I R, Kunz C & Willeke T, Themobot museum robot installations: A five year experiment, *Proc IEEE/RSJ Int Conf Int Rob Syst* (IEEE) 2003, 3636–3641, doi: 10.1109/IROS.2003.1249720.

4 Murtra A C, Tur J M & Sanfeliu A, Efficient active global localization for mobile robots operating in large and cooperative environments, *IEEE Int Conf Rob Autom* (IEEE) 2008, 2758–2763, doi: 10.1109/ROBOT. 2008.4543628.

5 Kaplan F, Talking AIBO: First experimentation of verbal interactions with an autonomous four-legged robot, In *Learn*

to behave: interacting agents CELE-TWENTE Work Lang Tech, **22** (2000).
6 Dautenhahn K, Woods S, Kaouri C, Walters M L, Koay K L & Werry I, What is a robot companion-friend, assistant or butler? *IEEE/RSJ Int Conf Intell Rob Syst* (IEEE) 2005, 1192–1197, doi: 10.1109/IROS.2005.1545189.
7 Pransky J, Social adjustments to a robotic future, *Wolf and Mallett*, (2004) 137–59.
8 Vaughan R, Sumpter N, Henderson J, Frost A & Cameron S, Experiments in automatic flock control, *Rob Auton Syst*, **31(1-2)** (2000) 109–117, doi: 10.1016/S0921-8890(99)00084-6.
9 Choudhury T, Aggarwal A & Tomar R, A deep learning approach to helmet detection for road safety, *J Sci Indus Res*, **79(6)** (2020) 509–512.
10 Feil-Seifer D & Mataric M J, Defining socially assistive robotics, *9th Int Conf Rehab Rob* (IEEE) 2005, 465–468, doi: 10.1109/ICORR.2005.1501143.
11 Hayashi K, Sakamoto D, Kanda T, Shiomi M, Koizumi S, Ishiguro H, Ogasawara T & Hagita N, Humanoid robots as a passive-social medium-a field experiment at a train station, *2nd ACM/IEEE Int Conf Human-Rob Interact* (IEEE) 2007, 137–144, doi: 10.1145/1228716.1228735.
12 Burgard W, Cremers A B, Fox D, Hähnel D, Lakemeyer G, Schulz D, Steiner W & Thrun S, The interactive museum tour-guide robot, *AAAI/IAAI*, (1998) 11–18, doi: 10.1016/S0004-3702(99)00070-3.
13 Sabanovic S, Michalowski M P & Simmons R, Robots in the wild: Observing human-robot social interaction outside the lab, *9th IEEE Int Work Adv Motion Control* (IEEE) 2006, 596–601, doi: 10.1109/AMC.2006.1631758.
14 Shiomi M, Kanda T, Glas D F, Satake S, Ishiguro H & Hagita N, Field trial of networked social robots in a shopping mall, *IEEE/RSJ Int Conf Intell Rob Syst* (IEEE) 2009, 2846–2853, doi: 10.1109/IROS.2009.5354242.
15 Sisbot E A, Alami R, Siméon T, Dautenhahn K, Walters M & Woods S, Navigation in the presence of humans, *5th IEEE-RAS Int Conf Humanoid Rob* (IEEE) 2005, 181–188, doi: 10.1109/ICHR.2005.1573565.
16 Shiomi M, Kanda T, Ishiguro H & Hagita N, A larger audience, please!—Encouraging people to listen to a guide robot, *5th ACM/IEEE Int Conf Human-Rob Interact* (IEEE) 2010, 31–38, doi: 10.1109/HRI.2010.5453270.
17 Shiomi M, Kanda T, Koizumi S, Ishiguro H &Hagita N, Group attention control for communication robots, *Int J Humanoid Rob*, **5(04)** (2008) 587–608, doi: 10.1142/S021984360800156X.
18 Reddy A K, Malviya V & Kala R, Social cues in the Autonomous Navigation of indoor mobile robots, *Int J Soc Rob*, **13** (2021) 1335–1358, doi: 10.1007/s12369-020-00721-1.
19 Martinez-Garcia E A & Akihisa O, Crowding and guiding groups of humans by teams of mobile robots, *In IEEE Work Adv Rob Soc Impacts* (IEEE) 2005, 91–96, doi: 10.1109/ARSO.2005.1511629.
20 Garrell A, Sanfeliu A & Moreno-Noguer F, Discrete time motion model for guiding people in urban areas using multiple robots, *IEEE/RSJ Int Conf Intell Rob Syst* (IEEE) 2009, 486–491, doi: 10.1109/IROS.2009.5354740.
21 Arulampalam M S, Maskell S, Gordon N & Clapp T, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, *IEEE Trans Signal Proc*, **50(2)** (2002) 174–188, doi: 10.1109/78.978374.
22 Everingham M & Zisserman A, Identifying individuals in video by combining'generative'and discriminative head models, *Tenth IEEE Int Conf Comput Vis* (IEEE) 2005, 1103–1110, doi: 10.1109/ICCV.2005.116.
23 Sivic J, Everingham M & Zisserman A, "Who are you?"-Learning person specific classifiers from video, *IEEE Conf Comput Vis Pattern Recog* (IEEE) 2009, 1145–1152, doi: 10.1109/CVPR.2009.5206513.
24 Tapaswi M, Bäuml M & Stiefelhagen R, "Knock! Knock! Who is it?" probabilistic person identification in TV-series, *IEEE Conf Comput Vis Pattern Recog* (IEEE) 2012, 2658–2665, doi: 10.1109/CVPR.2012.6247986.
25 Wolf L, Hassner T & Maoz I, Face recognition in unconstrained videos with matched background similarity, In *CVPR* (IEEE) 2011, 529–534, doi: 10.1109/CVPR.2011.5995566.
26 Barr J R, Bowyer K W, Flynn P J & Biswas S, Face recognition from video: A review, *Int J Pattern Recog Artif Intel*, **26(05)** (2012) 1266002, doi: 10.1142/S02180014 12660024.
27 Cinbis R G, Verbeek J & Schmid C, Unsupervised metric learning for face identification in TV video, *IEEE Int Conf Comput Vis* (IEEE) 2011, 1559–1566, doi: 10.1109/ICCV.2011.6126415.
28 Lu C & Tang X, Surpassing human-level face verification performance on LFW with Gaussian Face, *AAAI Conf Artif Intell*, **29(1)** (2015), doi: 10.48550/arXiv.1404.3840.
29 Sivic J, Everingham M & Zisserman A, Person spotting: Video shot retrieval for face sets, *Int Conf Image Video Retrieval* (Springer Berlin Heidelberg) 2005, 226–236, doi: 10.1007/11526346_26.
30 Kumar N, Kunju N, Kumar A & Sohi B S, Active marker based kinematic and spatio-temporal gait measurement system using LabVIEW vision, *J Sci Ind Res*, **69(8)** (2010) 600–605.
31 Smeulders A W, Worring M, Santini S, Gupta A & Jain R, Content-based image retrieval at the end of the early years, *IEEE Trans Pattern Analysis Mach Intell*, **22(12)** (2000) 1349–1380, doi: 10.1109/34.895972.
32 Malviya V & Kala R, Trackingvehicle and faces: Towards socialistic assessment of human behaviour, *IEEE Conf Inform Communication Technol* (IEEE) 2018, 1–6, doi: 10.1109/INFOCOMTECH.2018.8722427.
33 Azadi S, Moradi M & Esmaili A, Optimal balancing of PUMA-like robot in predefined path, *J Sci Indus Res*, **74(4)** (2015) 209–211.
34 Malviya V & Kala R, Socialistic 3D tracking of humans from a mobile robot for a 'human following robot' behaviour, *Robotica*, **41(5)** (2023) 1407–1435, doi: 10.1017/S0263574722001795.
35 Malviya V & Kala R, Trajectory prediction and tracking using a multi-behaviour social particle filter, *Appl Intell*, **52(7)** (2022) 7158–7200, doi: 10.1007/s10489-021-02286-6.
36 Malviya A & Kala R, Social robot motion planning using contextual distances observed from 3D human motion tracking, *Expert Syst Appl*, **184** (2021) 115515, doi: 10.1016/j.eswa.2021.115515.

37 Malviya A & Kala R, Learning-based simulation and modeling of unorganized chaining behavior using data generated from 3D human motion tracking, *Robotica*, **40(3)** (2022) 544–569, doi: 10.1017/S0263574721000679.

38 Malviya A & Kala R, Risk modeling of the overtaking behavior in the Indian traffic, *IEEE 25th Int Conf Intell Transp Syst* (IEEE) 2022, 2882–2887, doi: 10.1109/ITSC55140.2022.9922140.

39 Zhang K, Zhang Z, Li Z & Qiao Y, Joint face detection and alignment using multitask cascaded convolutional networks, *IEEE Signal Proc Lett*, **23(10)** (2016) 1499–1503, doi: 10.1109/LSP.2016.2603342.

40 Li H, Lin Z, Shen X, Brandt J & Hua G, A convolutional neural network cascade for face detection, *IEEE Conf Comput Vis Pattern Recog* (IEEE) 2015, 5325–5334, doi: 10.1109/CVPR.2015.7299170.

41 He K, Zhang X, Ren S & Sun J, Delving deep into rectifiers: Surpassing human-level performance on image net classification, *IEEE Int Conf Comput Vis* (IEEE) 2015, 1026–1034, doi: 10.1109/ICCV.2015.123.

42 https://www.youtube.com/watch?v=rABqTiqeA48