MiRo Person Recognition

Introduction

Person Recognition involves human target detection with face recognition, body recognition and tracking. Once the person is recognised the path finding and trajectory prediction is implemented for the robot to follow the person.

MiRo is an emotionally engaging fully autonomous robot that thinks and functions similar to animals. The behaviours and body language show the senses and decision-making processes of the robot. The MiRo has an SD card with Yocto Krogoth 2.1.1(Operating system) and Kinetic (ROS). The workstation supported is an Ubuntu 16.04 LTS (Operating system) and Kinetic (ROS). The MiRo has 7 sensors, 4 Processing/Comms and 9 actuators as shown in Figure 1 and 2. The neck has three Degrees of Freedom, both the ears rotate, tail droop and wag, eyelid open/close. All of these are provided with sensors. MiRo is suitable for developing companion robots.[8]

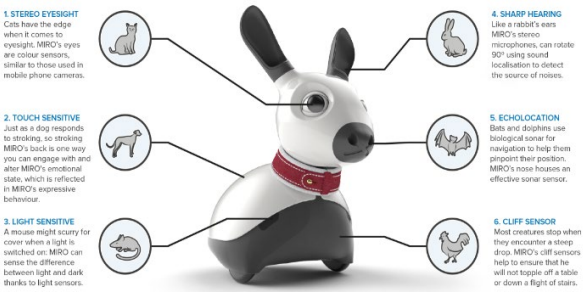


Figure 1

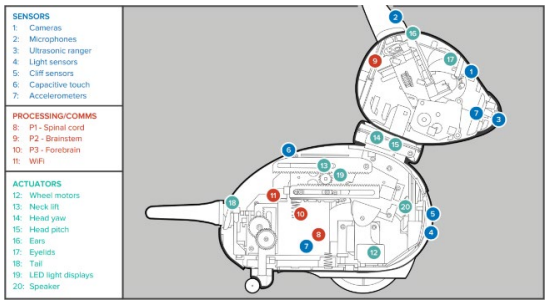


Figure 2

The Challenge faced in real world when integrating a robot into a family is the trust/familiarity factor and not be a major cultural shift from having a traditional pet. Hence the significance of the project undertaken is integrating a social robotic pet which mimics a behaviour of a real animal providing a psychological support such as identifying the house members and following them around.

Related Work

I. HOG detection:

The method employs image coordinates and depth value to store all the pixels from the depth image via point cloud. This technique comprises 4 steps, which includes shrinking the point cloud by filtration, resulting in an image with constant density. Moreover, the coordinates from the ground image are removed to cluster the point cloud data. In the next stage, a cluster is generated for each person in the frame [2]. The clusters are then classified using the HOG descriptor to determine the appearance and shape of the human. The example of how cluster will be created is shown in figure 3. With the histogram obtained (figure 4) and normalized, the classification of individuals is implemented by a supervised learning technique.

Figure 3

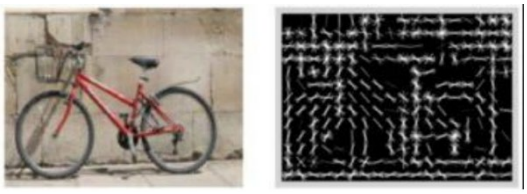


Figure 4

II. Human Pose Estimation feature-based tracking:

A method which can be used is person tracking based on human body pose transform tracking based on CNN based models. A skeleton of a person joint representation is overlayed on the detected image and is tracked and each joint position is updated each frame as part of the tracking [6]. For re-identification of the person while tracking a specific person, the trajectory of the points are estimated using trajectory replication-based techniques and the robot turns accordingly to bring the person back in the view.

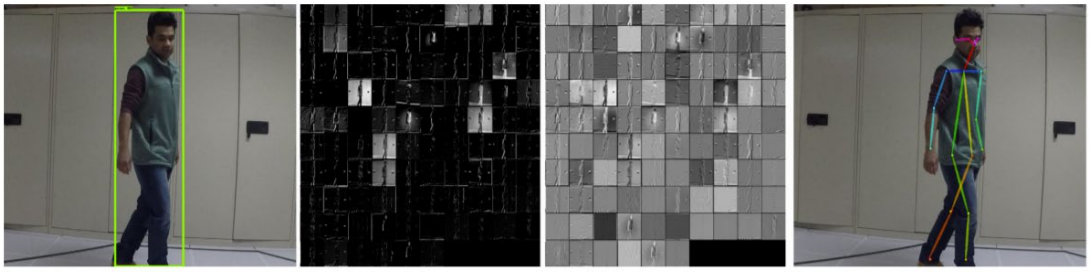


Figure 5

III. Rapid Object Detection:

The Rapid Object Detection algorithm is a quick and precise algorithm that is popularly known for detecting faces. It applies concepts such as AdaBoost for selecting the most beneficial features, Haar feature selection for edge and line detection, and uses 38 stages of Cascade classifier. The sliding window approach is implemented, and each sub-window image is reduced to 24x24 pixels length which is then converted to integral images and transferred to the cascade classifier. If the sub-window image produces a negative result at any stage, it is instantly dropped, otherwise, it goes to the next stage. If the sub-window is not discarded and clears all 38 stages, then a face is detected in that sub-window [12][14].

Methodologies

The Miro robot tracks the person after the recognition of the face and body, based on the architectural diagram (figure 6) the Miro robot tracks the person after the recognition of the face and body. Once tracked, the environment of the robot is mapped, and for each frame the robot will calculate the path from its position towards the person.

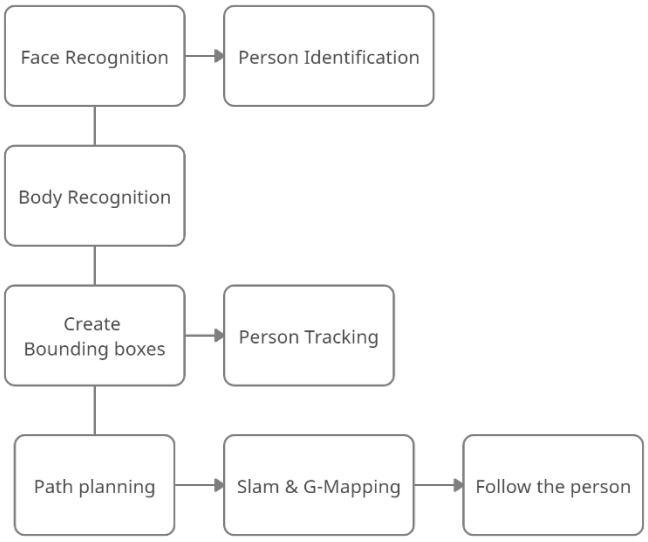


Figure 6

Face recognition:

1. **Using face_recognition package**:

i. Finding face:

The face_recognition package is being used to identify the different persons inside a given frame captured by the MiRo Bot. To identify a face, it generates a Histogram of Oriented Gradients (HOG) for a given frame and matches it with the HOG face pattern obtained from other training faces [11].

It then uses the face landmark estimation technique for generating 68 landmarks on the face, to perform operations such as rotation, scaling, and shearing the image for developing an image that resembles the face directly looking into the camera (figure 7).

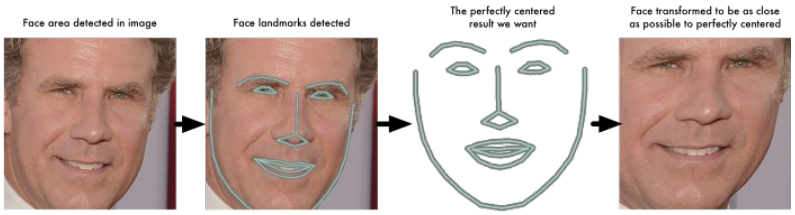


Figure 7

ii. Identifying the person:

Based on the extracted face, the package uses Deep Convolutional Neural Network to obtain 128 encoding points that are unique to each face. We obtain this encoding from the known face and match it with the unknown face encoding from the frame, and based on the closeness of both, we can determine whether a known person is in the frame or not [11].

2. **Using SIFT:**

The paper proposes a novel face recognition method using (Scale Invariant Feature Transform) SIFT and Bayesian approach which produces robust results in various environments including indoor with natural lighting [13].

i. Face Detection:

The Rapid Object Detection is used for extracting the face from the image which uses haar feature selection and Adaboost. The box is plotted on top of the detected face region.

ii. Face tracking:

A rectangular region is drawn on top of the box plotted in the previous step after increasing the length of each side by 2/3 of its original length.

iii. Face recognition:

a. Image pre-processing:

This step enhances the image by setting up the contrast and adjusting the illumination. For illumination, the image is divided into 16 regions and for each region, an average of the pixel value is taken to create a 4 x 4 grid. This grid is then bilinear interpolated to achieve the same size as the image and its complement is added to the primary image (figure 8).

Figure 8

Figure 9

b. SIFT feature extraction:

Rapid Object Detection is used again to identify the position of an eye. Form its position, the position of the other eye and nose and mouth region can be known. From these three regions, SIFT points are obtained and stored under a name given by the user (figure 9).

c. Finding the known person in a new frame:

The unknown person(f) in the frame can be identified by finding the level of similarity between its SIFT features and the stored features of known faces. Euclidean distance between the SIFT points is used to calculate the similarity and obtain a vector for it, $\vec{S} = \{s_1, s_2, \ldots, s_n\}$ (n being the total number of known faces). For any similarity in the vector, if it is greater than a threshold value, then the Bayesian approach is used.

Bayesian probabilities are used to determine the extent to which a face is similar. Two types of probabilities are calculated, relative and absolute. The absolute probability is calculated by finding the similarity between each known face and the given unknown face with respect to the sum of similarities between all faces (equation 1).

$$P_{abs}(s|f_i) = \frac{s_i}{\sum_{k=1}^{n} s_k}$$

Equation 1

The relative probability is calculated in a similar way. However, each similarity is divided by the number of SIFT (tp) points obtained for that similarity (equation 2).

$$P_{rel}(s|f_i) = \frac{\frac{s_i}{tp_i}}{\sum_{k=1}^{n} \frac{s_k}{tp_k}}$$

Equation 2

All the probabilities are stored and compared and if a probability is higher than the rest with a high margin, then the unknown person is identified.

Body Recognition:

When a robot overlooks a human face, body recognition is considered. The methodologies considered in this paper to detect humans entails Histogram of Oriented Gradients (HOG) based detectors and Online Adaboosting technique to categorize humans. These techniques necessitate the frame's depth image for detection.

I. Deriving Depth image:

The two cameras in MiRo robot renders the depth image of the frame via the following steps (1). The first step involves the elimination of tangential and radial distortions of the cameras, followed by adjusting the relative position between them to acquire a collinear image. Subsequently, from the two images, similar features are identified, and their discrepancies are computed. Thereby, the disparity map is obtained. The final step involves deriving the depth map from the equation 3, where f is the focal length, T is the baseline, x_l & x_r are the left & right camera image coordinates.

$$\frac{T-(x_l-x_r)}{Z-f} = \frac{T}{Z} \Rightarrow Z = \frac{f \cdot T}{x_l - x_r} \ .$$

Equation 3

II. Selected Online Adaboosting technique:

Human recognition and tracking are achieved by the Adaptive Boosting algorithm [3]. The classifier algorithm is initialised by two methods i.e., recognising the human within the predefined bounding box and compute the initial disparity of the human frame from the depth image, and improving the accuracy by eliminating 75% of the disparities. To compute the proportion of unwanted features, each depth pixel is predominant.

$$curDisp = Mean(I_p[I_p \in preDisp \pm \beta])$$

Equation 4

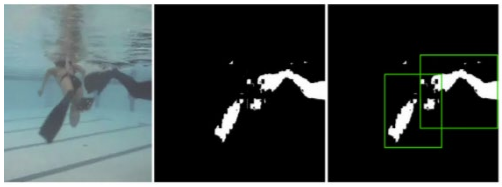
$$R = \frac{\sum [I_p \in preDisp \pm \beta]}{w * h}$$

Equation 5

On computing the initial disparity (preDisp), the disparity in the successive frames are estimated via Online AdaBoost algorithm which detects the positive patch. A threshold value is introduced in order to estimate the possible disparities of the human in a single patch image. The mean pixels of the attainable disparities(currDisp) are evaluated from the equation 4, which in turn is assigned to the preDisp. Thereby, the possible disparities are determined for every frame. The Depth Ratio R i.e., the minimum amount of unwanted features of the positive patch is derived from the equation 5. If R exceeds the threshold value, the classifier is updated with the current positive patch [7].

Person Tracking:

After the person is detected, a bounding box is put in the detected part of the image and is recognized as Region of Interest (ROI) for tracking. Two methodologies considered are colour histogram on ROI based tracking and contour-based approach [5]. The advantage of having a second tracking feature based on contour is due the possibility to multiple objects or persons wearing similar colour outfits.



Color-based tracking algorithms perform binary thresholding image, which is then refined to track divers.

Figure 10

Figure 11

Based on the normalized histogram generated for the region of interest, a shift in the peaks of the histogram in any given direction gives the information on where the person being followed will be in next frame.

In contour method, a mask is first generated at the start-up. And with each iteration, non-rigid features of the shape are tracked and updated in the model. With each passing frame after the person detection, a distance measure is calculated between the model generated and the edge detection of the image based on OpenCV canny edge detection. But contour-based tracking method is computationally expensive and hence used in combination of colour-based tracking.

Path Finding:

The final step for the robot is to follow the user around. The person tracking method with the left and right camera of the MiRo is responsible for tracing the person in the frame and aligning them in the centre of the view. The navigation package of ROS loads the dynamic map of the environment created with Gmapping. The path planner calculates the path from the robot to the person's position. The updated location position of the person is obtained for every frame, and the new path is calculated.[10]

The ability of the robot to reach the person's position can be considered with a tolerance of 0.1 meters. The MiRo's angular and linear velocity is set based on the distance and the steering angle required to reach the person's position. When an obstacle is identified by the sonar sensors, MiRo will start turning by 90 degrees while moving forward. This proceeds till the obstacle are not in sight.

i. **Trajectory prediction:**

The efficiency of the person-following robot is determined by its capacity to predict the person when they abruptly disappear from the view. This is possible by predicting the target's likely trajectory from the history with a regression model. A Support Vector Machine Regression (SVR) can provide a greedy fit on the nonlinear values as it uses the kernel functions.[9] Grid search is one of the simple ways to tune the parameters of the model. The target is tracked with person tracking. When the person is not visible in the view, the robot will search the target with pose, face or clothes.

Results

The Miro successfully recognized all known faces, and we are able to add new faces as well. After recognizing the face, the robot rotates its ear and moves towards the person. Due to the use of face_recgnition package, it was not able to find person in every frame and takes time to calculate the encoding values in the beginning. The package is also prone to illumination and works only if there is sufficient lighting provided.

The lights on the robot's body is turned on when it recognizes a person, but this affected its movement. After examining, it was identified that the messages from the light and movement were updated in the same topic, but of different behaviour. Consequently, the robot's movement messages are altered by the light messages.

Evaluation & Discussion

Face Recognition:

Face recognition using SIFT is robust in different illuminations and can learn a new face from a single frame. For a known person, the precision ranges from 96.65% to 100% and recall vary from 33% to 57.32%. For an unknown person, the precision is above 90%. The model takes 3 seconds to identify a person on a system with Pentium D, 1 GB Ram, and a 2.8 GHz processor. The model doesn't always find a person in a given frame.

i. **Comparison of SIFT model with the current model:** The current model which uses the face_recognition package is prone to illumination and doesn't necessarily find a person in each frame. Based on the observation, if the threshold is too low then it discards the correct person and if it is too high then it misclassifies the known person.

Body Recognition:

The robot will recognize the person irrespective of their pose or same clothes worn. It will also be able to identify the person in different poses. Since the algorithm is not specific to identifying humans, the robot will be able to recognize any of the objects (eg: Bags, clothes, etc). The figure 13 explains with the evaluation results of different scenarios [7].

The classification model can be replaced, since the Adaboosting algorithm is similar to Random Forest classifier which might result in a frail and less robust algorithm. The robot is more likely to get confused or bemused with slight changes in the bounding boxes [7].

Person Tracking using Colour histogram and Contour features:

Processing time is 0.1 secs while running the tracking code standalone and when used with all modules included, it has capability to process 7 frames per second. The minimum computation capability required is 200MHz with 128MB of RAM which is far lower than the onboard hardware available on MiRo.

The contour-based approach tackles the room and body illumination issues as it does not rely on homogenous illumination as long as there is colour differentiation between person and the location behind.

The main weakness of using a colour-based approach is failure in detection when the person is wearing black or white clothes in a similarly coloured background.

Path finding:

The difference between the obstacle and person is identified clearly. Hence, it moves away from the obstacle and towards the person. In path finding, when people start walking at a higher speed than that of the robot there is a possibility of losing track of the target. The transition of tracking from one person to another might not be smooth.

Conclusion

This report describes a novel way that can make MiRo recognise and follow the person. Based on the analysis of chosen methodologies, it was identified that the following approaches can be used in order to solve the problem. Rapid Object Detection for finding the face in different illumination settings. Selected Adaboosting algorithm for recognizing the body of the person. The colour histogram and contour features technique provided by the open pose, which is an open-source library. The G-Mapping is used to create a dynamic map and the trajectory prediction is used if the person is not in the frame. The path planner will give the path from the robot's position to the target person avoiding the obstacles.

References

1. Ling, Fuhai & Jimenez-Rodriguez, Alejandro & Prescott, Tony. (2019). Obstacle Avoidance Using Stereo Vision and Deep Reinforcement Learning in an Animal-like Robot. 71-76. 10.1109/ROBIO49542.2019.8961639. Available at https://eprints.whiterose.ac.uk/158352/1/Ling%202019%20ROBIO%20Obstacle%20Avoidance.pdf
2. Caroline Queva (2013). Human Following Behavior for an Autonomous Mobile Robot. Degree project in Computer Science Second cycle. Stockholm, Sweden 2013. https://www.diva-portal.org/smash/get/diva2:703048/FULLTEXT01.pdf.
3. BenMauss (Mar 2021), "*Adaboost in Image Classification*", https://levelup.gitconnected.com/adaboost-in-image-classification-8dcc1799e53d
4. Grabner, Helmut & Grabner, Michael & Bischof, Horst. (2006). Real-Time Tracking via On-line Boosting. Proceedings of British Machine Vision Conference (BMVC). 1. 47-56. 10.5244/C.20.6. https://www.researchgate.net/publication/221259753_Real-Time_Tracking_via_On-line_Boosting
5. Schlegel, Christian et al, Vision Based Person Tracking with a Mobile Robot. https://www.researchgate.net/publication/221259161_Vision_Based_Person_Tracking_with_a_Mobile_Robot
6. Md Islam et al, Person-following by autonomous robots: A categorical overview. https://journals-sagepub-com.ezproxy.lib.rmit.edu.au/doi/pdf/10.1177/0278364919881683
7. Chen, Bao Xin & Sahdev, Raghavender & Tsotsos, John. (2017). Person Following Robot Using Selected Online Ada-Boosting with Stereo Camera. 48-55. 10.1109/CRV.2017.55. https://www.researchgate.net/publication/323063775_Person_Following_Robot_Using_Selected_Online_Ada-Boosting_with_Stereo_Camera
8. Consequential Robotics Ltd, "Miro Beta Developer Kit", http://consequentialrobotics.com/miro-beta, 2020. [Online; accessed 18-June-2021]
9. M. Kim, M. Arduengo, N. Walker, Y. Jiang, J. W. Hart, P. Stone, and L. Sentis,"An architecture for person-following using active target search," 2018.
10. Y. Nagumo and A. Ohya, "Human following behavior of an autonomous mobile robot using light-emitting device," *Proceedings 10th IEEE International Workshop on Robot and Human Interactive Communication. ROMAN 2001 (Cat. No.01TH8591)*, 2001, pp. 225-230, doi: 10.1109/ROMAN.2001.981906.
11. Geitgey, A 2016, *Machine Learning is Fun! Part 4: Modern Face Recognition with Deep Learning*, Medium, blog, viewed 20 June 2021, https://medium.com/@ageitgey/machine-learning-is-fun-part-4-modern-face-recognition-with-deep-learning-c3cffc121d78
12. Lee, S 2020, *Understanding Face Detection with the Viola-Jones Object Detection Framework*, Medium, blog post, viewed 20 June 2021, https://medium.com/@ageitgey/machine-learning-is-fun-part-4-modern-face-recognition-with-deep-learning-c3cffc121d78
13. C. Cruz, L. E. Sucar and E. F. Morales, "Real-time face recognition for human-robot interaction," 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition, 2008, pp. 1-6, doi: 10.1109/AFGR.2008.4813386.
14. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, 2001, pp. I-I, doi: 10.1109/CVPR.2001.990517.
15. Wikipedia 2021, Viola–Jones object detection framework, Wikipedia, viewed 20 June 2021, https://en.wikipedia.org/wiki/Viola%E2%80%93Jones_object_detection_framework

Appendix

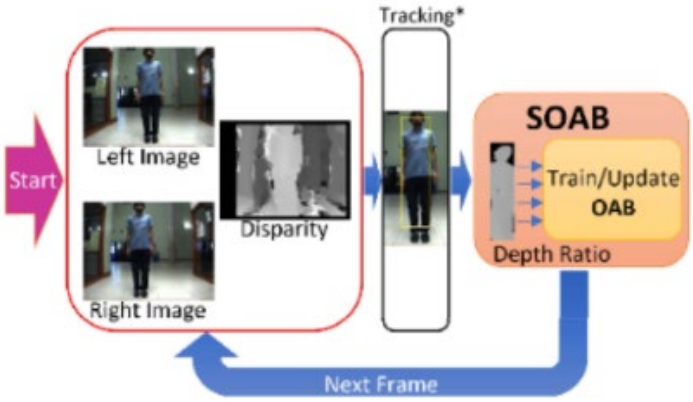


Figure 120

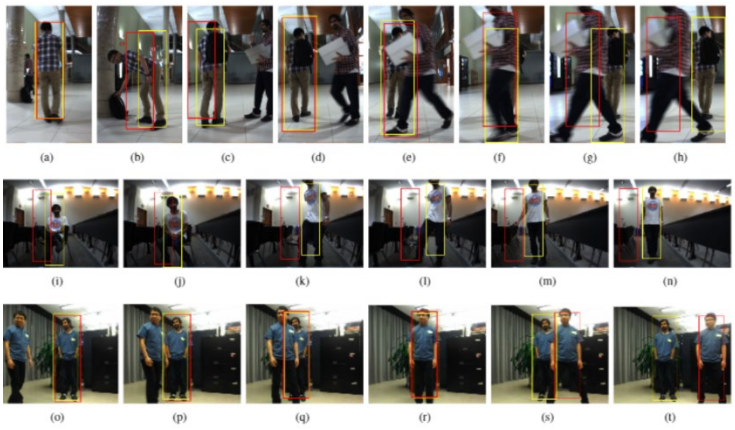


Figure 13