On-Line learning-based Fuzzy Clustering Modelling in Real-Time Object Tracking

Chang Wui Lee and Ong Sing Ling

School of Computing and Creative Media (SCM), University of Technology Sarawak, 96000 Sibu, Sarawak.

Abstract: Real-time object tracking has been one of the challenging issues in autonomous navigated robotic design. Most of the solutions proposed the off-line calculation or powerful computation unit on-board, where collection of data samples is provided in prior. While a complete data collection is hard to achieve practically, this paper proposed an on-line learning-based fuzzy adaptive resonance theory (Fuzzy ART) with Takagi-Sugeno-Kang Fuzzy model aimed to cluster and predict object positions from live-stream video frames. In this study, an object is extracted from a video frame, and object' centroid position is fed for fast-learning. Fast-learning mode of Fuzzy ART is used for real-time clustering the positions. While the position obtained are considered fuzzy, the clusters formed unique patterns at each cluster and represent them as fuzzy sets. The fuzzy sets are used to predict the object position using the Takagi-Sugeno-Kang Fuzzy Inference System (TSK-FIS) model. The proposed algorithm contributes to an on-line and unsupervised fuzzy modelling for object tracking to predict the object's position in real-time. In order to validate the practicality of the proposed approach, two experiments are conducted, i.e., using simulated data samples, and a case study using real-time hand movement tracking. A bounding box for region of interest on a frame is moved according to the predicted positions, while the robustness of the tracking is measured through the box's movements. The proposed model is self-improving over time, which has great advantage for self-correction in autonomous applications.

Keywords: Fuzzy adaptive resonance theory, on-line fuzzy modelling, clustering.

INTRODUCTION

Object detection and tracking is the key technique in the field of intelligence transportation [1]–[5] where camera technologies are implemented as the main input sensor for neural network-based modelling. For example, an edge traffic flow detection scheme [1] uses cameras to detect vehicles, at the traffic light junctions across the city, and neural network-based deep learning algorithm is used to model and predict the vehicles' population. In an Unmanned Ariel Vehicle (UAV) application [2], a camera frame is extracted using the Oriented FAST and Rotated Binary Robust Independent Elementary Features (ORB) algorithm, while the K-Nearest Neighbour (KNN) is used to match the image descriptors for the object's position prediction. In two wheeled robot cooperatively carrying an object in unknown environments [3], cameras are used as the input for the zero-order Takagi-Sugeno Fuzzy model, that is evolutionarily optimised using the Particle Swarm Optimisation (PSO), to predict and correct track for the robots.

Machine learning has been gaining the popularity in most applications in object detection and tracking due to the availability of high performance of the processing devices [6]. Off-line learning methods require both large data samples and extensive resources to train the data samples while optimising the objective function of a model for an application. In real application, the data samples are not static, where the collected data samples might be deviated from their original results due to the changes at the real environment, i.e., concept drifting [7]. In this case, the trained model might be not applicable when the environment is different. Data-driven and on-line learning approaches has been proposed to overcome concept drifting problem [7]–[9] with consideration of the stream data samples are generated in real-time from the enclosed environment, as well as, unexplored environment. On-line learning learn a data sample in a single-pass learning manner [10], where a data sample is learn and discarded after that one after another. Thus, reduced the RAM memory usage during the training process.

Clustering has been popular for its unsupervised learning feature to group unlabeled data samples into a number of clusters of unique features [11]. Data-driven applications in clustering have been reported in [12], [13]. The resultant clusters reveal the hidden message/information of the data samples [14]–[17] that is often being interpreted as expert knowledge, which is typically represented as fuzzy rules [18]. Takagi-

Sugeno-Kang Fuzzy Inference System (TSK-FIS) [19] models the qualitative aspect of human knowledge and reasoning processes without precise quantitative analyses using fuzzy rules. Clustering in TSK-FIS applications have been reported in [19]–[22], where clusters are used to determine the number of rules and TSK-FIS predicts the corresponding outputs using heterogenous regression functions.

In this paper, we proposed an on-line learning based fuzzy clustering modelling in real-time object tracking using Fuzzy Adaptive Resonance Theory (Fuzzy ART) clustering algorithm for two objectives, which are, (1) to cluster similar object's positions into groups for knowledge representation, and to improve the prediction of the object's positions in real-time. A stream of video frames is pre-processed using non-robust image processing for object detection, and a data sample (center of gravity), i.e., the object's position, is obtained for the learning in the Fuzzy ART algorithm to group the data sample into the existing cluster or a new cluster. The single-pass learning manner is implemented by the following assumptions: (1) the cluster structure is stable after each learning; (2) the cluster's size is expanding to adapt the data sample into the cluster; and (3) the clustering structure is expanding by increasing the number of clusters. At the clustering structure, fuzzy rules are extracted, and the membership functions are generated at each cluster for knowledge representation to the data stream. Zero-order TSK-FIS model is used to predict the object's position at current time.

The organisation of the paper is as follows. In Section 2, the preliminary of studies in the Fuzzy ART clustering algorithm and TSK-FIS modelling are presented. The proposed methodology is explained in Section 3, and the experimental results are discussed in Section 4. Lastly, a conclusion is remarked in Section 5.

PRELIMINARY OF STUDY

In this section, the Fuzzy ART clustering algorithm and zero-order TSK-FIS model procedure are explained. Let denotes an input data sample as vector $x = [x_1, ..., x_m] \in [0,1]$, where m is the number of features (dimension), for the Fuzzy ART and TSK-FIS model.

Fuzzy Adaptive Resonance Theory (Fuzzy ART)

The learning of Fuzzy ART algorithm is divided into three levels, i.e., data level, category matching level, and category updating level. The following notations are used. A cluster is associated with a prototype weight vector $w_i = [w_{i,1}, ..., w_{i,2m}] \in [0,1]$. Three user-defined parameters are predetermined at the algorithm, which are, choice parameter, i.e., denoted as

$\beta \in (0, \infty)$, learning rate, i.e., denoted as $\alpha \in [0,1]$, and vigilance value, i.e., denoted as $\rho \in [0,1]$.

The algorithm starts when a data sample is obtained at the data level. Normalisation of the data sample is conducted by adding a complement coding to the data vector. Let denotes the normalised data sample as $A = [x_1, ..., x_m, x_1^c, ..., x_m^c] \in [0,1]$, where $x_*^c = 1 - x_*$ is the complement value and the number of dimensions of A is $2m$. A is then fed for clustering in category level. Let assume it is an empty cluster, where no previous learned cluster available. At this level, the cluster is initialised with a prototype weight of $w_1 = [1, ..., 1]$. Equation (1) is used to compute the matching degree of A to w_i across all prototype weights, where $A \wedge w_i$ is the minimum operator, i.e., $\min(A, w_i)$, $|w_i|$ is the summation operator, i.e., $\sum_{j=1}^{2m} w_{i,j}$. The prototype weight with the maximum $T(i)$ is the winner.

$$T(i) = \frac{|A \wedge w_i|}{\beta + |w_i|} \in [0,1] \qquad (1)$$

The winner is carried forward to the category updating level. At this level, a vigilance test is conducted, where Equation (2) is applied.

$$\frac{|A \wedge w_i|}{|A|} \geq \rho \qquad (2)$$

If the vigilance test is satisfied, the prototype weight is called committed, and the weight is updated using Equation (3).

$$w_i = \alpha(A \wedge w_i) + (1 - \alpha)w_i \qquad (3)$$

If Equation (2) is not satisfied, the winner is uncommitted, and repeat Equation (1) without the uncommitted prototype weight vectors. A new winner is selected, and vigilance test is conducted at the new winner. This process is repeated until a winner is committed. Otherwise, create a new cluster with a prototype weight vector of $w_{k+1} = A$, where k is the number of existing clusters. Finally, the prototype weight vectors are stored for the next data sample fetched.

Takagi-Sugeno-Kang FIS (TSK-FIS) Model

The TSK-FIS model discussed in this section is focused on the multiple input multiple output (MIMO) link system, where the m dimensional data sample is treated as m inputs to predict m outputs.

Three steps in TSK-FIS model, i.e., fuzzification, inference engine, and defuzzification, are considered in

this section. The following notations are used. Fuzzy sets at j-th dimension is denoted as $B^j = \left[B_1^j, ..., B_{k_j}^j\right]$, where $k_j \in [0, \infty)$ is the number of fuzzy sets at j-th dimension. Let consider a bell-shaped membership function at each fuzzy set as shown in Figure 1. A membership degree is denoted as $\mu^j \in [0,1]$ that describe the degree of belonging of a x_j value to a fuzzy set.
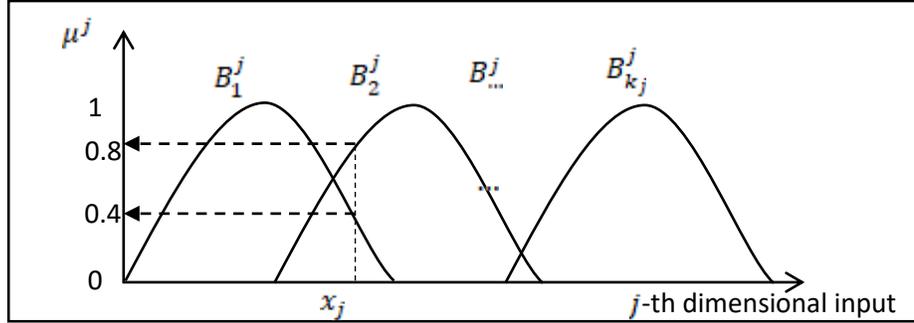

Figure 1: Example of k_j bell shaped membership functions at j-th dimension.

For example, a j-th dimension input value x_j is considered. The membership values to $B_{1,...,k_j}^j$ is 0.8, 0.4, 0, …, 0 to indicate the degree of belonging of x_j to B_1^j is 0.8, degree of belonging of x_j to B_2^j is 0.4, and zeros at $B_{...}^j$ and $B_{k_j}^j$, respectively.

At the fuzzification step, a crisp value of x_j is fuzzified into k_j dimensional membership values, i.e., $x_j \rightarrow \left[\mu_1^j, ..., \mu_{k_j}^j\right]$. The membership value is typically computed using Gaussian function, where Equation (4) is applied, where $v_{B_i^j} \in [0,1]$ is the mean value of i-th fuzzy set (i.e., B_i^j), and $\sigma_{B_i^j} \in [0,1]$ is the gaussian width at B_i^j. The larger the $\sigma_{B_i^j}$, a wider separation of a fuzzy set is formed.

$$\mu_i^j = e^{\frac{-\left(x_j - v_{B_i^j}\right)^2}{2\sigma_{B_i^j}^2}} \qquad (4)$$

The fuzzified input is fed for the inference engine step, where fuzzy IF-THEN rule is applied. Equation (5) is an example of i-th rule in j-th dimension, i.e., denoted as rule r_i^j, and $y(r_i^j)$ is a predefined heterogenous consequence function based on the application.

$$rule\ r_i^j : IF\ x_j\ is\ B_i^j\ THEN\ y(r_i^j) \qquad (5)$$

Typically, the number of rules is determined by the experts in the field, and it is common to have multiple rules triggered at the same time.

At defuzzification step, it is an aggregation method to aggregate all $y(r_i^j)$ triggered to form a crisp output value. Equation (6) is used to obtain the predicted result y^j at j-th dimension.

$$y^j = \frac{\sum_{i=1}^{k_j} \mu_i^j y(r_i^j)}{\sum_{i=1}^{k_j} \mu_i^j} \qquad (6)$$

When a rule r_i^j is triggered, $\mu_i^j > 0$. Otherwise, $\mu_i^j = 0$.

PROPOSED METHODOLOGY

In this section, the proposed on-line learning-based fuzzy clustering modelling in real-time object tracking is explained with Figure 2.

The methodology starts when a RGB (red, green, blue) video frame is captured from a stream of live video frames. A squared region of interested, sized 20% of the frame width, is positioned at the center of the frame to reduce the computation at object detection. An object is initially positioned inside the squared region of interest to ease object detection. The region of interest is filtered with image processing methods, i.e., (i) transforming RGB format to grey scaled image, (ii) removing the lower and upper intensity pixels using a predetermine threshold, (iii) converting the grey scaled image into binary image. An object is detected at a region of white pixels from the binary image, and a center of gravity is computed from the region of white pixels. Thus, the initial position of the object is detected.

The obtained position is fed for clustering to group the similar object positions for knowledge learning in single-pass learning manner, where Fuzzy ART clustering algorithm from Section 2 is revisited

with the learning rate set to one, i.e., $\alpha = 1$. A number of prototype weight vectors is obtained from the Fuzzy ART learning. Information of the prototype weight vectors is extracted by locating the lower and upper boundaries of the clusters, which is determined by Equation (7), where v_j is the center point of i-th cluster in j-th dimension.

$$v_{B_i^j} = \frac{1 - w_{i,m+j} + w_{i,j}}{2} \qquad (7)$$

Equation (4) is revisited, where $\sigma_{B_i^j} = \left(1 - w_{i,m+j} - w_{i,j}\right)/2$, to compute the gaussian membership functions. Thus, $k_j = k$, where the number of fuzzy sets at j-th dimension is set to the number of clusters.

The number of inference rules are obtained from the clusters, where rule r_i^j is derived as follow.

$$rule\ r_i^j : IF\ x_j\ is\ B_i^j\ THEN\ y\left(r_i^j\right) = v_{B_i^j}$$

A video frame is captured.

A squared region of interest is positioned at the center of the frame.

The frame is filtered for object detection.

A center of gravity of the object is obtained, i.e., denoted as $\boldsymbol{x}$.

$\boldsymbol{x}$ is clustered using Fuzzy ART clustering algorithm.

Gaussian membership functions are obtained at each cluster.

Inference rules are obtained from the clusters' prototype weight vectors.

Expected object position is located using the TSK-FIS aggregation function.

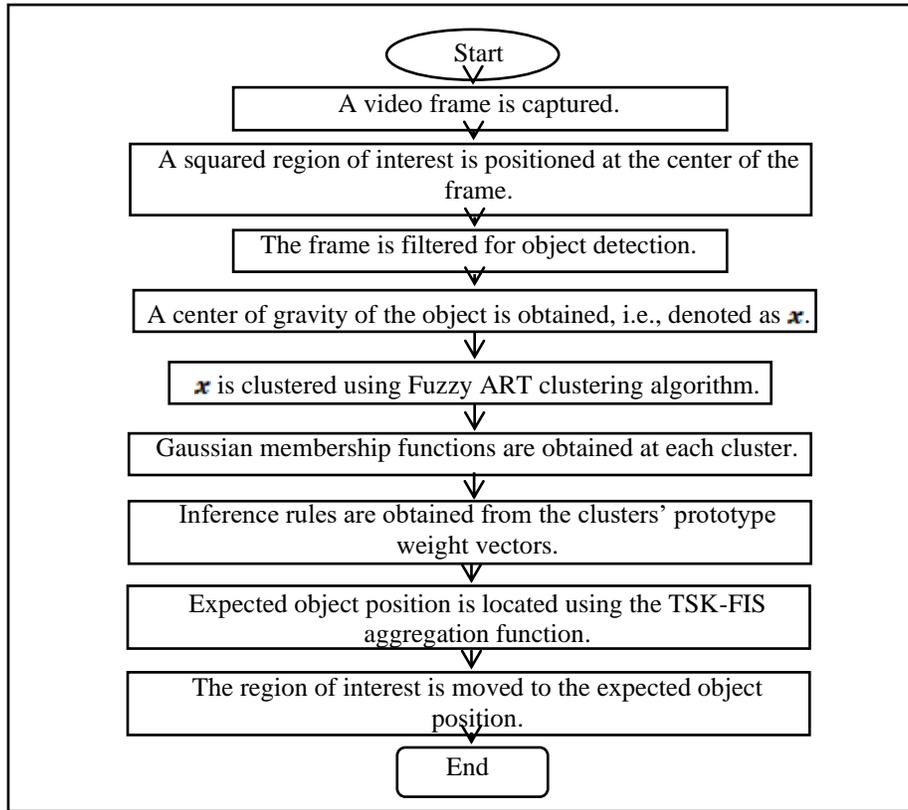The region of interest is moved to the expected object position.

Figure 2: Flow chart of on-line learning-based fuzzy clustering modelling in real-time object tracking for one video frame

Equation (6) is revisited for the TSK-FIS aggregation method to predict the expected object's position on the frame.

The region of interest is then moved to the expected object's position obtained, where the center of the region is set to the object's position on the video frame at current time. The process is repeated for the next captured frame continuously until the video has stopped.

RESULTS AND DISCUSSION

This section, the experiments are divided into two parts, i.e., (i) experiment using simulated data, where on-line learning is analysed, and (ii) experiment using case study, where the practicality issue is investigated.

Simulated data

The simulated data are generated using a Sine function in degree unit. In order to mimic the non-robustness at the data sample, random noises of maximum 40% noise deviation rate to $1 + \sin(i)$, $i = 0, \dots, 360$. Equation (8) depicts the stream of data samples used in this

experiment, while Figure 3 illustrate the graphical plot of the data samples.

$$x = \left[\frac{i}{360} \, , noise(1 + \sin(i)) \right] \qquad (8)$$
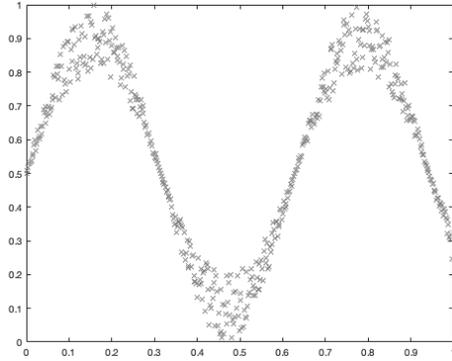

Figure 3: The simulated data samples with maximum 40% deviation rate from the original Sine function

(a) Rules extraction from clusters

Figure 4 depicts the clusters structure using the Fuzzy ART algorithm, where the boundaries of each cluster is depicted using a rectangular-shaped box and the data samples of each cluster is illustrated using a colour map to ease visual inspection to the recognition rate at each cluster. From the figure, all the data samples are successfully recognised by the clusters, where all the data samples are positioned within the boundaries of the clusters. It is noted that some boundaries are overlapping each other, arrow pointed in Figure 8. This happens when the concept drifting occurred to trained data samples, where the data samples that are previously belong to cluster A is now "drifted" to cluster B.
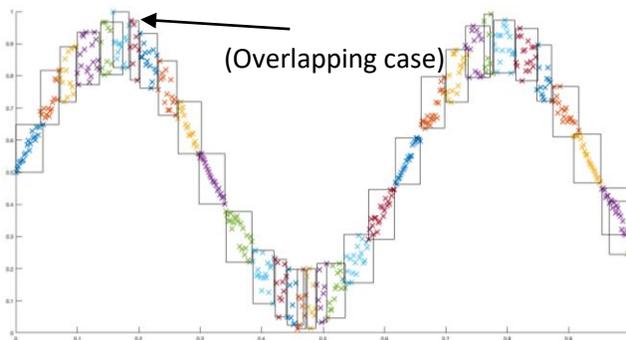

(Overlapping case)

Figure 4: Clustering structure of Fuzzy ART. The input data samples and boundaries are visualised with colours based on their clusters

In Figure 5, the information from the clusters are interpreted using Gaussian membership function at each feature of the clusters, where the x-axis of Figure 3 is the feature 1 and y-axis of Figure 3 is the feature 2. From Figure 5, feature 1 has a more consistent gaussian width across the clusters, due to the no-repeated values of the x-axis in Figure 3. For the membership functions constructed at the repeated value of the y-axis in Figure 3, Figure 5(b) shows some of the gaussian membership functions are overlapping or similar at the clusters.

(b) Real-time tracking with proposed methodology

The membership functions from Figure 5 is applied for TSK-FIS model to predict the output positions at x-axis and y-axis using Figure 5(a) and Figure 5(b), respectively. Figure 6 shows the outputs from the learning before concept drift, i.e., Figure 6(a), and after concept drift, i.e., Figure 6(b). The accuracy of the prediction is measured using the squared-box as explained in Section 3 (methodology), where the prediction is assumed correct when the prediction box covers the data sample.

The learning of data samples one after another (single-pass manner) at the initial stage, especially when the first new data sample is fed, the model is called lack of knowledge representation, and the prediction is not accurate. This phenomenon is clearly shown in Figure 6(a) pointed with a dark arrow. From the figure, the wrong prediction position centred at the red box's boundaries are obviously not bounded the responding data sample. However, as the next new data samples are
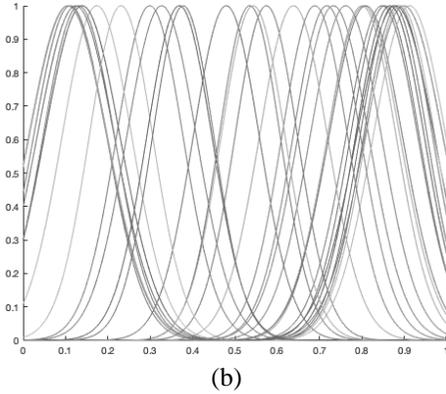

(a)

(b)
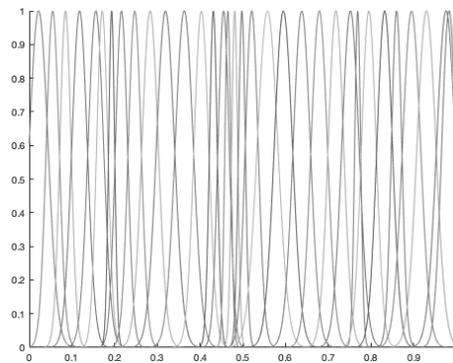
Figure 5: Membership functions of (a) x-position, and (b) y-position

fed for learning, the knowledge representation has improved. The Fuzzy ART algorithm captured the concept drifting, where cluster of belong of a data sample is shifted to another cluster, and the membership functions further improved the knowledge representations of the clusters. Thus, the accuracy of the prediction is significantly improved when a learned data sample is for prediction with the improved knowledge representation.

Figure 7 plotted the positions of the original data samples, on-line prediction (before concept drift), and test prediction (after concept drift). From the figure, the prediction with noisy data is improving using on-line prediction, and the prediction accuracy is further optimised after a number of data samples are learnt. From this experiment, on-line learning for predicting an output position of a data sample shows improving performance as more data samples are learnt (more experience is gained).

Case study

The experiment is further analysed with a case study in a real-time hand movement tracking using a webcam, as shown in Figure 8(a). A current video frame is captured after each learning. A region of interest is used to crop out from the capture frame for image processing. The cropped image is filtered simply by subtracting the red elements and green elements of the pixels, by assuming the background has less red pixel elements as compared to the object, as shown in Figure 8(b).
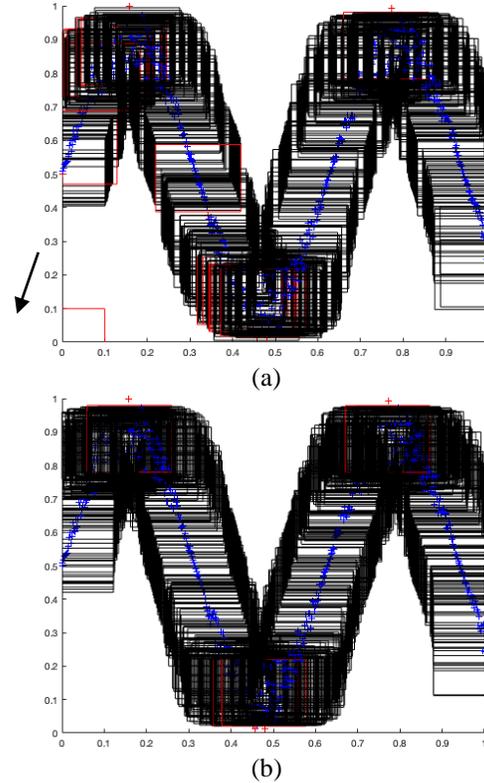

(a)

(b)

Figure 6: Data with 40% random noise, with black box indicating true prediction while red box indicating false prediction; (a) 17 false predictions at new data, (b) 4 false prediction at trained data samples
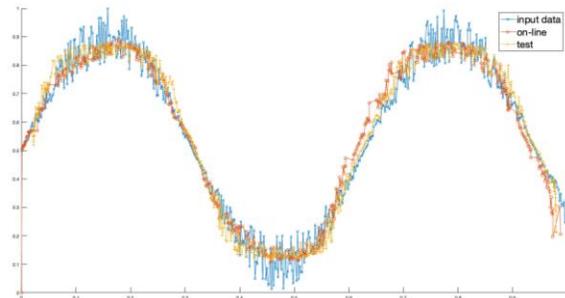

Figure 7: Comparison of the predicted positions using the data with 40% random noise in blue and red plot indicates the prediction result on first trial, while the yellow plot is the prediction on the next trials.

The object is detected successfully after converting into binary image to highlight the white pixels as the object, where the center of gravity is then computed from the largest white region of the binary image, shown in Figure 8(c). The center of gravity is marked as red cross symbol on Figure 8(b). It is noted that the data sample is the centre of gravity detected that is positioned on the captured video frame.
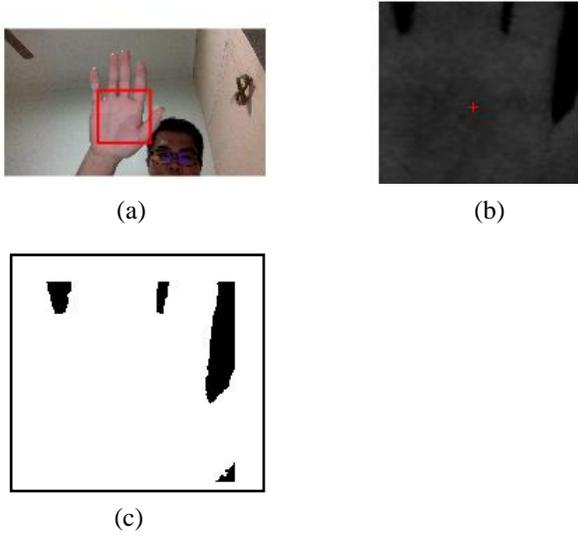
(a)

(b)

(c)

Figure 8: Hand tracking using webcam camera. (a) the region of interest of red bounding box. (b) subtraction of red and green elements. (c) the binary segmentation at threshold 0.078

Figure 9 shows the cluster structure generated from the detect hand positions (data samples) in scale of [0,1] to ease analysis. There are nine clusters formed, where all the data sample are recognised and stayed within the cluster boundaries. The gaussian membership functions generated at the end of the tracking is shown in Figure 10, and the predictions results are plotted on Figure 11 to show the path of the hand movement and tracking results. From Figure 11, an arrow is pointed at the input data (blue coloured), which has shown that the objected is initially not being detected due to the camera setting that starts with an empty pixel on the first frame captured. The prediction result has not being affected by the object at the first frame.

From the results findings, the tracking is improving when the proposed methodology is exposed to more new data samples. In other words, the more the application continued, the more accurate the prediction results.
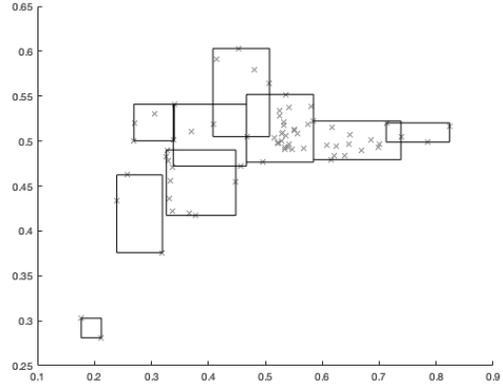

Figure 9: Clustering structure of the detected object's positions generated from live video frame.
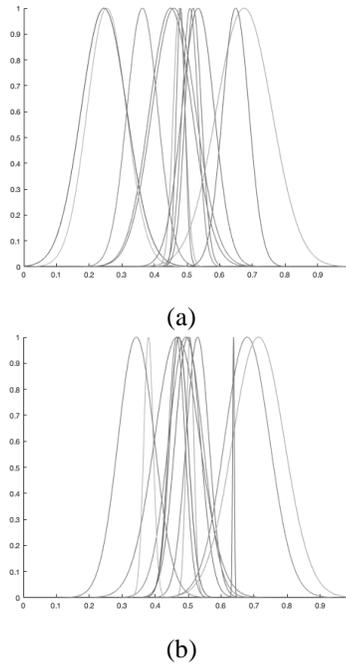

(a)

(b)

Figure 10: Membership functions at clusters. (a) the membership functions at x-feature, and (b) the membership functions y-feature.
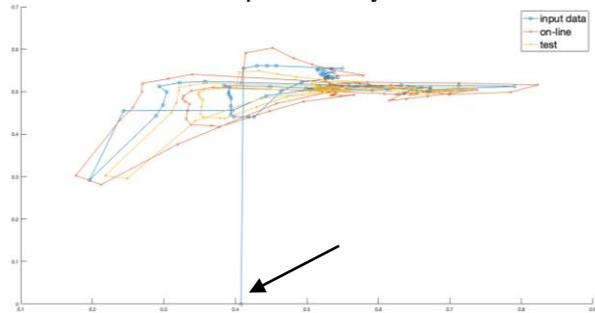

Figure 11: Prediction results of the hand movements with hand positions plotted with blue, red indicates first trial predictions, and yellow indicates the second trials predictions

CONCLUSION

On-line learning-based fuzzy clustering modelling in real-time object tracking methodology is proposed using the Fuzzy ART clustering algorithm and TSK-FIS model to predict the real position of the object's position at a time. The Fuzzy ART clustering with the learning rate set to 1's learnt the stream of object positions in a single-pass manner, and the clustering structure is updated after each learning of a new data sample. Gaussian membership functions are then generated from the resultant clustering structure at each learning to formed a knowledge representation of the current data samples. The knowledge representation then used by TSK-FIS model to predict the expected object's position based on the history results. Experiments results are obtained from a simulation data and a case study. The simulation result with noise data samples shown the capability of the proposed method to cluster noisy data and the knowledge representation with Gaussian membership functions depicts the deviation of the noise data samples at each cluster as faithfully. A case study has been successfully implemented to track hand movement on a webcam camera. The prediction results are depicted with the x- and y- coordinates as the center of a bounding box, where the object is expected to be at. From the findings, the prediction results are improved over time when more data samples are fed for training.

The x- and y- coordinates at the object positions produce a multi-input-multi-output (MIMO) link system, which is typically used in autonomous applications to control the robotic movements at the x- and y-axes, and contribute to the future direction of this work.

ACKNOWLEDGEMENT

REFERENCES

[1] Chen, C., Liu, B., Wan, S., Qiao, P. and Pei, Q. 2021. An edge traffic flow detection scheme based on deep learning in an intelligent transportation system. IEEE Trans. Intell. Transp. Syst., 22 (3), 1840–1852.

[2] Chen, P., Dang, Y., Liang, R., Zhu, W., and Xiaofei, H. 2017. Real-time object tracking on a drone with multi-inertial sensing data. IEEE Trans. Intell. Transp. Syst., 19(1), 131–139.

[3] Juang, C. F., Lai, M. G. and Zeng, W. T. 2014. Evolutionary fuzzy control and navigation for two wheeled robots cooperatively carrying an object in unknown environments. IEEE Trans. Cybern., 45 (9), 1731–1743.

[4] Chandan, G., Jain, A. and Jain, H. 2018. Real time object detection and tracking using deep learning and openCV. In 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), 1305–1308.

[5] Hossain, S. and Lee, D. 2019. Deep learning-based real-time multiple-object detection and tracking from aerial imagery via a flying robot with GPU-based embedded devices. Sensors, 19(15), 3371.

[6] Elhoseny, M. 2020. Multi-object detection and tracking (MODT) Machine learning model for real-time video surveillance systems. Circuits, Syst. Signal Process, 39(2), 611–630.

[7] Lu, J., Liu, A., Song, Y. L. and Zhang, G. Q. 2020. Data-driven decision support under concept drift in streamed big data. Complex Intell. Syst., 6(1), 187–163.

[8] Chao, L., Tan, C., Wang, X. and Zheng, Y. 2019. An evolving recurrent interval type-2 intuitionistic fuzzy neural network for online learning and time series prediction. Appl. Soft Comput. J., 78, 150–163.

[9] Nallaperuma, D. et. al. 2019. Online incremental machine learning platform for big data-driven smart traffic management. IEEE Trans. Intell. Transp. Syst., 20(12), 4679–4690.

[10] Lughofer, E. 2011. Evolving fuzzy systems - Methodologies, advanced concepts and applications.

[11] Xu, R. 2005. Survey of clustering algorithms for MANET. IEEE Trans. Neural Networks, 16(3), 645–678.

[12] Tian, Y., Zheng, R., Liang, Z., Li, S., Wu, F. and Li, M. 2021. A data-driven clustering recommendation method for single-cell RNA-sequencing data. Tsinghua Sci. Technol., 26(5), 772–789.

[13] Geva, A. et. al. 2021. EClinicalMedicine data-driven clustering identi fi es features distinguishing multisystem in fl ammatory syndrome from acute COVID-19 in children and adolescents. EClinicalMedicine, 40, 101112.

[14] Chang, W. L., Pang, L. M. and Tay, K. M. 2017. Application of self-organizing map to failure modes and effects analysis methodology. Neurocomputing, 249, 314–320.

[15] Chang, W. L. and Tay, K. M. 2014. A new evolving tree for text document clustering and visualization. 223.

[16] Lin, Z,. Wen, F. , Ding, Y. and Xue, Y. 2017. Data-driven coherency identification for generators based on spectral clustering. IEEE Trans. Ind. Informatics, 14(3), 1275–1285.

[17] Krishna, K. S. and Seal, A. 2020. Clustering

analysis using an adaptive fused distance. Eng. Appl. Artif. Intell., 96, 103928.

[18] Li, J. Yang, L. Qu, Y. and Sexton, G. 2018. An extended Takagi – Sugeno – Kang inference system (TSK +) with fuzzy interpolation and its rule base generation. Soft Comput., 22(10), 3155–3170.

[19] Tzafestas, S. G. and Zikidis, K. C. 2001. NeuroFAST: On-line neuro-fuzzy ART-based structure and parameter learning TSK model. IEEE Trans. Syst. Man, Cybern. Part B, 31(5), 797–802.

[20] Jeng, J. 2003. Fuzzy regression clustering algorithm for TSK fuzzy. In Proceedings 2003 IEEE International Symposium on Computational Intelligence in Robotics and Automation. Computational Intelligence in Robotics and Automation for the New Millennium (Cat. No. 03EX694), 201–206.

[21] Adabelief, P. et. al. 2021. FCM-RDpA: TSK fuzzy regression model construction using fuzzy C-means clustering, regularization, DropRule, and Powerball AdaBelief. Inf. Sci. (Ny), 574, 490–504.

[22] Mousavi, S., Esfahanipour, A. and Zarandi, M. H. F. 2021. A modular Takagi-Sugeno-Kang (TSK) system based on a modi ed hybrid soft clustering for stock selection. Sci. Iran, 28(4), 2342–2360.