An Interval Type-2 Fuzzy Logic System for Human Silhouette Extraction in Dynamic Environments

Bo Yao[1], Hani Hagras[1], Danniyal Al Ghazzawi[2], and Mohammed J. Alhaddad[2]

[1] School of Computer Science and Electronic Engineering, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK
[2] King Abdulaziz University, Jeddah, Saudi Arabia

Abstract. In this paper, we present a type-2 fuzzy logic based system for robustly extracting the human silhouette which is a fundamental and important procedure for advanced video processing applications, such as pedestrian tracking, human activity analysis and event detection. The presented interval type-2 fuzzy logic system is able to detach moving objects from extracted human silhouette in dynamic environments. Our real-world experimental results demonstrate that the proposed interval type-2 fuzzy logic system works effectively and efficiently for moving objects detachment where the type-2 approach outperforms the type-1 fuzzy system while significantly reducing the misclassification when compared to the type-1 fuzzy system.

Keywords : Interval type 2 fuzzy logic, Silhouette extraction, Human tracking.

1 INTRODUCTION

Accurate human silhouette (or outline) segmentation from a video sequence is important and fundamental for advanced video applications such as pedestrian tracking and recognition, human activity analysis and event detection. Advanced human detection and identification approaches like [1], [2] can be utilized for silhouette extraction. However, such methods are commonly of high computational complexity and hence not suitable for dynamic and complex environments. Hence, there is a need for silhouette extraction methods which are computationally efficient and that are able to operate in dynamic and complex environments.

The first step for silhouette extraction is background modeling and subtraction to detect moving targets as foreground objects. In [3], an approach based on a single Gaussian modal was developed which employed a simple robust method to handle moving objects and slow illumination changes. However, there are several limitations in this method such as learning stage necessity for background distribution, and robustness deficiency for situations like sudden illumination changes, slow moving objects, etc. To address these problems, the Gaussian Mixture Model (GMM) was proposed [4]. In this model, each pixel is modeled using n Gaussian distributions. GMM is effective to overcome the shortcomings of single Gaussian model and hence GMM is extensively recognized as a robust approach for background modeling and subtraction. Therefore, in this paper, GMM is utilized for foreground detection. However, it is unreasonable to simply consider GMM foreground as human silhouette in real-life environments because there are numerous noise factors and uncertainties to handle which include:

- Varying light condition
- Reflections and shadows
- Moving objects attached to human silhouette (a book, a chair, etc.).

To handle these problems and detach the moving objects from the human silhouette, a type-1 Fuzzy Logic System (T1FLS) was proposed [5]. This T1FLS is capable of handling to an extent the uncertainties mentioned above, however, the extracted silhouette will be degraded due to misclassification of the proposed T1FLS. Hence, in this paper, we will present an Interval Type-2 Fuzzy Logic System (IT2FLS) which will be able to handle the high uncertainty levels present in real-world dynamic environments while also reducing the misclassification of extracted silhouette. The IT2FLS used similar type-1 membership function as the ones presented in [5] as principal membership functions which are then blurred to produce the type-2 fuzzy sets used in this paper. We have also used the same rule base as [5] to allow for a fair comparison with the results reported in [5].

In this proposed system, GMM is adopted for original foreground detection, then a IT2FLS is performed to detach the moving objects from the human silhouette. We have performed several real-world experiments where it was shown that the proposed IT2FLS is effective to reduce the misclassification and the quality of the extracted human silhouette is much improved when compared to the T1FLS.

The rest of this paper is organized as follows. In section 2, we provide a brief overview of type-2 FLSs. Section 3 presents the proposed IT2FLS. Section 4 presents the experiments and results and finally the conclusions and future work are presented in section 5.

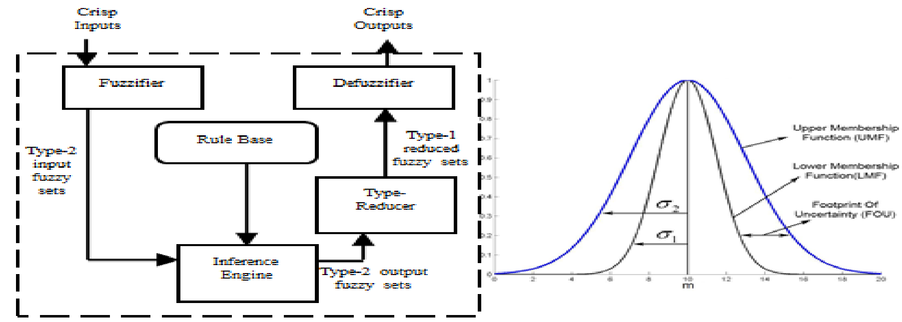


Fig.1.(a) Structure of the type-2 FLS. (b) An interval type-2 fuzzy set

2 A BRIEF OVERVIEW OF THE IT2FLS

The IT2FLS depicted in Fig. 1a) [6] uses interval type-2 fuzzy sets (such as the type-2 fuzzy set shown in Fig. 1b) [6] to represent the inputs and/or outputs of the FLS. In the interval type-2 fuzzy sets all the third dimension values are equal to one [6], [7].

The use of interval type-2 FLS helps to simplify the computation (as opposed to the general type-2 FLS) [8].

The interval type-2 FLS works as follows [6], [7], [8]: the crisp inputs from the input sensors are first fuzzified into input type-2 fuzzy sets; singleton fuzzification is usually used in interval type-2 FLS applications due to its simplicity and suitability for embedded processors and real time applications. The input type-2 fuzzy sets then activate the inference engine and the rule base to produce output type-2 fuzzy sets. The type-2 FLS rule base remains the same as for the type-1 FLS but its Membership Functions (MFs) are represented by interval type-2 fuzzy sets instead of type-1 fuzzy sets. The inference engine combines the fired rules and gives a mapping from input type-2 fuzzy sets to output type-2 fuzzy sets. The type-2 fuzzy output sets of the inference engine are then processed by the type-reducer which combines the output sets and performs a centroid calculation which leads to type-1 fuzzy sets called the type-reduced sets. There are different types of type-reduction methods. In this paper we will be using the Centre of Sets type-reduction as it has reasonable computational complexity that lies between the computationally expensive centroid type-reduction and the simple height and modified height type-reductions which have problems when only one rule fires [6], [7]. After the type-reduction process, the type-reduced sets are defuzzified (by taking the average of the type-reduced set) to obtain crisp outputs that are sent to the actuators. More information about the interval type-2 FLS and its benefits can be found in [6], [7], [8].

3 THE PROPOSED IT2FLS FOR HUMAN SILHOUETTE EXTRACTION

Fig. 2 provides an overview of proposed IT2FLS approach for human silhouette extraction. In the video capturing stage, source images are captured using a stationary camera. The images are then analyzed using GMM to detect foreground. After that, the foreground detected by GMM is partitioned into small $n \times n$ blocks, for example 2×2 blocks. Then Human tracking is performed by global nearest neighbor (GNN) [9] to obtain the human centroids. Based on the partitioned foreground and the obtained human centroids, human silhouette extraction is carried out.

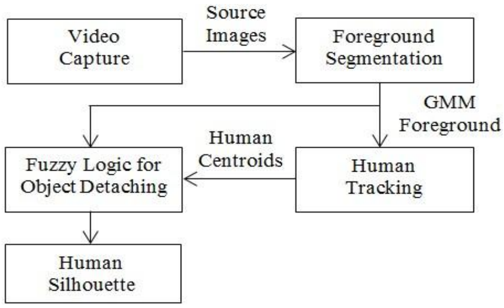


Fig. 2. An overview of the proposed system.

Indeed, advanced human detection and identification approaches can be utilized for silhouette extraction. However, those methods are commonly of high computational complexity and not robust for dynamic and complex environments. Hence in [5] a

T1FLS was presented and in this paper we will present an IT2FLS for silhouette extraction.

Suppose that we are working on frame *i*, and the foreground image of frame *i* obtained by GMM is denoted by O_i, and the silhouettes in frame *i-1* that have been properly segmented are denoted by O_{i-1}. As described above, the foreground in O_i may contain the human body and moving non-human objects attached to silhouette. To detach the moving objects and refine the human silhouette, the fuzzy logic system is developed based on the following observations:

1. If an image block in O_i belongs to the human body, it is of high probability to match an image block in O_{i-1} in a good match degree. SAD (sum of absolute difference) between their corresponding blocks in frame *i* and frame *i-1*, is used to measure the matching degree between the image block in O_i and its best match block in O_{i-1}.
2. If the distance between this block and human centroid is far, the probability that this block belongs to the human body is low.
3. If the amount of its neighbor blocks with high probability belonging to the human body is huge, for example having good matches in O_{i-1}, or having low distance to human centroids etc., then, the probability of this block also belongs to the human body is high.

Based on observations above, the following variables of each block are calculated.

- SAD of motion estimation. For every image block in O_i, its best match image block in frame *i-1* is searched. And the matching degree is used to describe its SAD variable.
- The distance between this block and the human centroid.
- The amount of its neighborhood with high probability of belonging to human body, for example, have a good match block in human body, low centroid distance, etc.

The rules of the type-2 fuzzy system should remain the same as the T1FLS in [5] and we will use the similar type-1 fuzzy sets in [5] as the principal membership functions which are then blurred by 10% (the 10 % was determined empirically to balance between robustness to noise and system performance) to produce the type-2 fuzzy sets in our IT2FLS. The membership functions for the inputs and output of the IT2FLS are shown in Fig 3. The rule base of the IT2FLS is the same as [5] and it is as follows:

1. If SAD is very low AND Neighborhood is huge AND Distance is close,
   THEN Silhouette is high.
2. If SAD is large AND Neighborhood is small AND Distance is very far,
   THEN Silhouette is low.
3. If SAD is low AND Neighborhood is large AND Distance is medium,
   THEN Silhouette is high.
4. If SAD is medium AND Neighborhood is medium AND Distance is medium,
   THEN Silhouette is medium.
5. If SAD is large AND Neighborhood is medium AND Distance is medium,
   THEN Silhouette is medium.
6. If SAD is large AND Neighborhood is large AND Distance is close,
   THEN Silhouette is high.

7. If SAD is medium AND Neighborhood is large AND Distance is medium, THEN Silhouette is high.
8. If SAD is medium AND Neighborhood is large AND Distance is close, THEN Silhouette is high.
9. If SAD is very large AND Neighborhood is small AND Distance is far, THEN Silhouette is low.
10. If SAD is medium AND Neighborhood is small AND Distance is very far, THEN Silhouette is low.
11. If SAD is low AND Neighborhood is huge AND Distance is medium, THEN Silhouette is high.
12. If SAD is low AND Neighborhood is medium AND Distance is medium, THEN Silhouette is high.

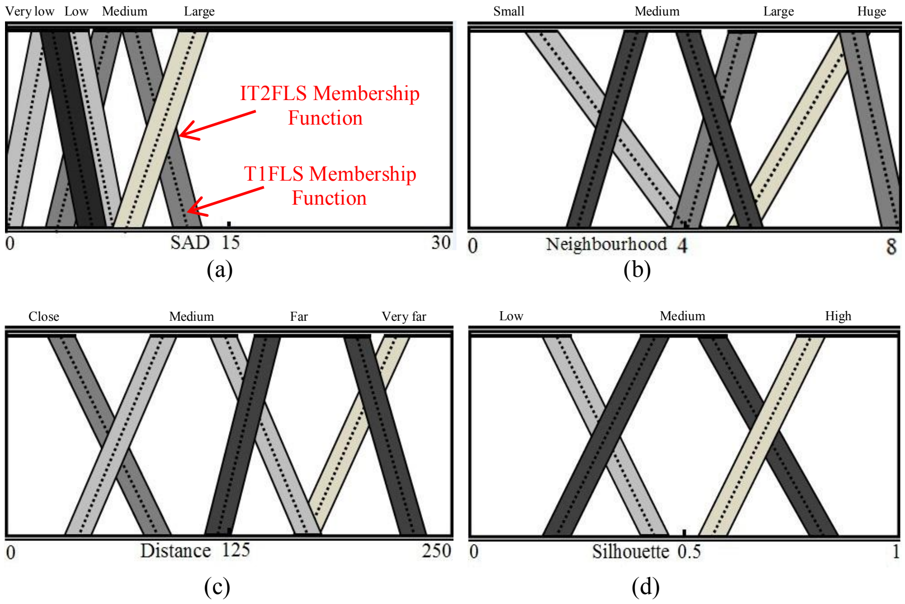


Fig. 3. The interval type-2 fuzzy sets employed in the inputs of our IT2FLS for (a) SAD, (b) Neighborhood, (c) Distance, (d) The Silhouette output of the IT2FLS.

4 EXPERIMENTAL RESULTS

We have performed several real-world experiments to validate our proposed approach and to compare the performance of the IT2FLS and the T1FLS presented in [5]. The ground truth data is captured from cameras deployed around our laboratory of smart living room to analyze people's regular activity. The aim of the experiments was to validate that the proposed IT2 fuzzy system is effective to detach the moving objects from human silhouette with much fewer misclassifications than T1 fuzzy system.

As shown in Fig. 4 to 8, column (a) shows the source images; column (b) shows the original foreground detected by GMM; column (c) illustrates the results after the T1FLS; column (d) exhibits the results after using IT2FLS. In our case, human silhouette is represented by pixels having a higher degree than the 0.5 degree of the fuzzy silhouette.

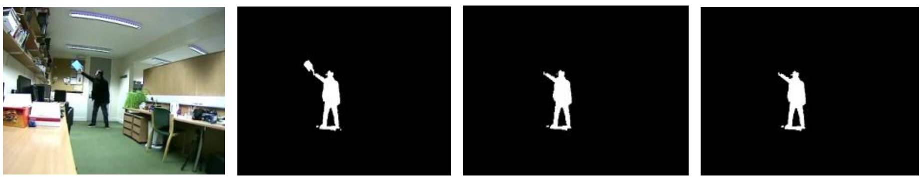
In Fig. 4, the results of "Raising a book" demonstrate that the book attached to human is eliminated after using the fuzzy based systems. The experiment shows that two fuzzy systems extract a proper silhouette and that they are able to detach the book from human body. However, as mentioned above, the silhouette extracted by T1FLS is degraded due to misclassification (as confirmed by [5]) while IT2FLS can address this problem and reduce the misclassification.

In Fig. 5, enlarged silhouette images of "Raising a book" are provided. As can be seen in Fig. 5(b), the edge the silhouette of T1FLS, (highlighted with red rectangles) is degraded to coarser edge due to misclassification when compared to the original foreground. However, as displayed in Fig. 5(c), the IT2FLS achieves a same human silhouette as the original foreground while the book is detached.

To demonstrate the robustness of the proposed system, more experiments have been done in various environments. In Fig. 6, results in an outdoor environment of single pedestrian are shown, via this experiment, we can see that the proposed system is working effectively in an outdoor environment. In Fig. 7, it can be seen in the images that the reflection of human body which is a noise factor is detected as foreground by GMM while the reflection is eliminated by utilizing fuzzy logic. In Fig. 8, results in an outdoor environment crowded with people are shown. In this complex outdoor environment with more noises and uncertainties, the proposed system demonstrates the robustness by extracting human silhouette with a promising result. In these experiments, for the purpose of comparison, Table 1 provides the average and standard deviations for the misclassification and accuracy for T1 and IT2 fuzzy systems. In Table 1, it can be seen clearly that the misclassification of the proposed IT2FLS is reduced significantly compared to T1FLS while the IT2FLS results also in higher accuracy than the T1FLS.

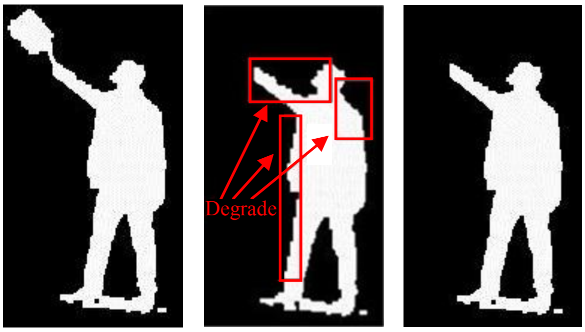
Experiment Name	T1 Average Misclassify	T1 STDEV Misclassify	T1 Average Accuracy	IT2 Average Misclassify	IT2 STDEV Misclassify	IT2 Average Accuracy	Frame Used
Single person	125.74 pixels	41.19 pixels	93.23%	0.61 pixels	1.37 pixels	99.9791%	383
Multi-person with reflection	221.38 pixels	74.73 pixels	95.23%	0.54 pixels	1.39 pixels	99.9951%	246
Crowded environment	637.62 pixels	180.59 pixels	90.21%	12.84 pixels	16.46 pixels	99.9004%	255

Table 1. Comparison of misclassification and accuracy.



(a) (b) (c) (d)

Fig. 4. The experiment of "Raising a book"; (a) source images, (b) foreground detected by GMM, (c) extracted silhouette after using T1FLS, (d) extracted silhouette after using IT2FLS.

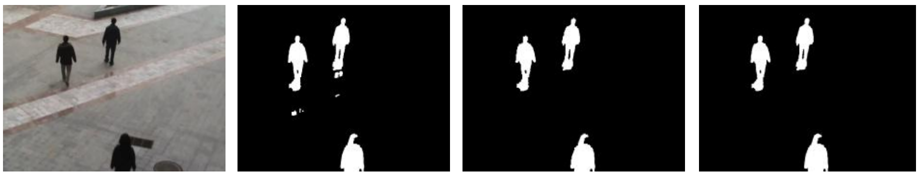


(a) (b) (c)

Fig. 5. Enlarged images of "Raising a book"; (a) foreground detected by GMM, (b) extracted silhouette after using T1FLS, (c) extracted silhouette after using IT2FLS.

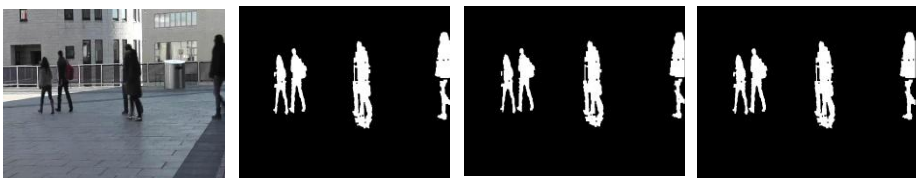
(a) (b) (c) (d)

Fig. 6. The experiment of "Single person".



(a) (b) (c) (d)

Fig. 7. The experiment of "Multi-person with reflection".



(a) (b) (c) (d)

Fig. 8. The experiment of "Crowded environment".

5 CONCLUSIONS AND FUTURE WORK

In this paper, we have presented an IT2FLS for improved silhouette extraction in dynamic real-world environments. Due to the huge complexity of the dynamic and real-life environment, the problem of detaching moving objects from human silhouette gets quite complicated. To address this problem without high computational complexity, we firstly use GMM to detect foreground, then an IT2FLS is employed for moving objects detachment. We have conducted several real-world experiments which have shown that the proposed IT2FLS is effective to detach objects and the misclassification is greatly reduced compared to a similar T1FLS while the IT2FLS results also in high accuracy for silhouette extraction compared to the T1FLS.Hence, by utilizing IT2FLS, the proposed system obtain silhouette extraction with good robustness to noise factors and uncertainties such as light condition changes, reflection of human body, and moving objects attached to the human silhouette, etc., in dynamic indoor/outdoor environments.

For our future ongoing research, we intend to extend the proposed algorithm for automatic learning which will enable the system to be more robust in dynamic environments. We also aim to apply the proposed system in high-level vision applications such as event detection, human activity recognition, etc.

REFERENCES

[1] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” Conference on Computer Vision and Pattern Recognition (CVPR), 2005.

[2] Y. Wu, T. Yu, “A field model for human detection and tracking, ” IEEE Transactions on Pattern Analysis and Machine Intelligence 28(5) (2006) 753–765.

[3] Aubert, G. and Kornprobst, P., “Mathematical Problems in Image Processing. Partial Differential Equations and the Calculus of Variations,” Applied Mathematical Sciences, vol. 147, Springer, New York, 2001.

[4] B. Katz, J. Lin, C. Stauffer, E. Grimson, “Answering questions about moving objects in surveillance videos, ” AAAI Spring Symposium on New Directions in Question Answer, 2003.

[5] X. Chen, Z. He, J. Keller, D. Anderson, and M. Skubic, “Adaptive Silhouette Extraction in Dynamic Environments Using Fuzzy Logic,” Proc., WCCI, Vancouver, BC, July, 2006.

[6] J. Mendel, Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions, Upper Saddle River, NJ: Prentice-Hall, 2001.

[7] H. Hagras, “A Type-2 fuzzy controller for autonomous mobile robots”, Proc. of the 2004 IEEE international conference on fuzzy systems, Budapest Hu, 2004, pp. 965-970.

[8] Q. Liang and J. Mendel, “Interval type-2 fuzzy logic systems: theory and design”, IEEE Trans. on Fuzzy Systems, Vol 8. pp 535-550, Oct. 2000.

[9] S. Blackman, R. Popoli, Design and Analysis of Modern Tracking Systems. Artech, 1999.