Mathware & Soft Computing 12 (2005) 185-198

Applying Fuzzy Logic in Video Surveillance Systems

J.M. Molina[1], J. García[1], O. Pérez[1], J. Carbó[1],
A. Berlanga[1] and J.I. Portillo[2]
[1]Univ. Carlos III de Madrid. Depto. de Informatica
Avda de la Universidad Carlos III, 22
Colmenarejo 28270. Spain
[2] Univ. Politecnica de Madrid. E.T.S.I. Telecomunicacion
Ciudad Universitaria s/n.
Madrid 28040. Spain

Abstract

In this work, the application of fuzzy logic in surveillance systems based on cameras is analyzed. Three different fuzzy systems have been tested and compared with a crisp decision system. The first one has been developed using an expert knowledge, the second one was learned from recorded videos, and a third one is developed as a refinement taking into account evaluation with ground truth. In all cases, the core of the system is the association function, in which the developed fuzzy system takes decision about what blobs (detected pixels grouped in a zone) belong to what tracks. In this work the surveillance video system is deployed in an airport. It is embedded in an A-SMGCS Surveillance function for airport surface, based on video data processing, in charge of the automatic detection, identification and tracking of all interesting targets (aircraft and relevant ground vehicles). The system evaluation has been developed using an evaluation function specifically designed for this type of problem. Results obtained with real data in representative ground operations show different capabilities for each system to solve complex scenarios and to improve tracking accuracy.

1 Introduction

A minimal requirement for automatic video surveillance system in industrial application [21] is the capability of tracking multiple objects or groups of objects in real conditions. A typical video surveillance system is composed of several processes: (1) A predictive process of the image background, usually Gaussian models are applied to estimate variation in the background; (2) a detector process of moving targets, detector process works over the previous and actual acquired frames.; (3) a grouping pixel process, this process groups correlates adjacent detected pixels to conform detected regions; (4) an association process, this process evaluate which detected blob should be considering as belonging to each existing target; and (5) a tracking system that maintains a track for each existing target.

Surveillance system depends on many parameters that should be adjusted for a specific implementation. The core of this process is the evaluation of surveillance results. The main point is the definition of a metric to measure the quality of a proposed set of configuration parameters [6]. There are many works [19] [20] that evaluate video surveillance systems against the ground truth or with synthetic images. In this work we extract the truth values from real images and they are stored in a file. In this file each target is located and positioned in each frame. Targets in the file are defined by six attributes: number of frame, track identifier, min and max value in coordinates x and y of the rectangle that surrounds the target.

In this work the surveillance video system is deployed in an airport. The application of video technology in airport areas in a new way to support ground traffic management inside the Advanced Surface Movement, Guidance and Control Systems (A-SMGCS) [11] [10] [1]. The video system used in this work is based on a previously developed prototype, intended to analyze the integration of video technology in A-SMGCS Surveillance function for Madrid/Barajas Airport. This work has been developed jointly by GRPSS group (Grupo de Procesado de Seal y Simulacin from Universidad Politcnica de Madrid) and GIAA group (Grupo de Inteligencia Artificial Aplicada from University Carlos III de Madrid). Specifications and details of this video system have appeared in several publications [3] [4] [5]. The specifications for A-SMGCS require the identification and accurate tracking of all aircraft and vehicles in the airport movement area, in order to improve awareness of surface traffic, conflict monitoring and guidance in a wide range of weather conditions. Basically, camera sensors are being explored as an alternative for surveillance in this area, used as a complementary source of data to conventional sensors such as surface movement radars [22].

The system architecture is a coupled tracking system where the detected objects are processed to initiate and maintain tracks representing the real targets in the scenario and estimate their location and cinematic state. The tracking feedback over detector allows coherent system behaviour and solves specific problems in this application such as "ghost" targets. The system captures the frames in the video sequence and uses them to compute background estimation. Background statistics are used to detect contrasting pixels corresponding to moving objects. These detected pixels are connected later to form image regions referred to as blobs. Blobs are defined with their spatial borders, generally a rectangular box, centroid location and area. Then, the tracker re-connects these blobs to segment all targets from background and track their motion, applying association and filtering processes.

The association process assigns one or several blobs to each track, while not associated blobs are used to initiate tracks. Map information and masks are used to tune specific aspects such as detection, track initiation, update parameters, etc. It is in the association logic where differences in the systems evaluated appear. Three different fuzzy systems have been tested and compared with a hard decision system. Rules for the first one were obtained using expert knowledge, while those for the second were learned from recorded videos. The rules for the third one have been developed as a refinement taking into account evaluation with ground truth.

In the next section, the proposed metric is presented. In section third, the fuzzy systems proposed are presented. Systems output in several scenarios are presented in section four, indicating the response for complex situations, with real image sequences of representative ground operations. Finally, some conclusions are presented.

2 Evaluation System

One of the most important aims of our study is to calculate some parameters which allow the evaluation of the performance of our tracking system. To achieve this goal, the measurements given for the tracking system are compared with the ideal output. This ground truth is the result of a careful study from pre-recorded video sequences and a subsequent process in which a human operator annotates the images, marking boxes bounding each target.

This process can be explained for each video in several steps as follows:

1. The most interesting objectives are selected in order to analyze their trajectories. The criterion for selection is the size and position of the different targets in the videos. The bigger target the better, and the more difficult to be distinguished from a close object the more interesting for our study.

2. The coordinates of the targets are selected frame by frame by surrounding them with rectangles and taking the upper left corner and lower right corner as location of our objectives at this moment. This location is referred to the upper left corner of the complete image which represents the pixel (0, 0). Then, the range of values varies from 0 to 767 pixels in the x-axis and from 0 to 575 in y-axis. Thus, the ground truth can be defined as a set of rectangles that define the trajectory of each target.

3. Finally, the ground truth data are stored in a table which will be used to compare to the result tracks of the tracking system.

The results trajectories have to be as similar as possible the ground truth tracks. Thus, the next step is the comparison of the ideal trajectories with the detected ones so that a group of parameters can be obtained to analyze the results and determine the quality of our detections.

2.1 Ideal trajectories

As stated above, each target is located and positioned in each frame and these data are recorded in a table of seven columns:

1. Number of frame which has been analyzed. There are so many lines a frame as selected targets in this specific frame.

2. Track identifier, which will be a number between 1 and the number of tracks, given to a specific trajectory.
3. Value of the minimum x coordinates of the rectangle which surrounds the target.
4. Value of the maximum x coordinates of the rectangle which surrounds the target.
5. Value of the minimum y coordinates of the rectangle which surrounds the target.
6. Value of the maximum y coordinates of the rectangle which surrounds the target.
7. Number of line in the text file.

Figure 1 below shows the format of the data for the ideal values stored after a careful extraction process which is done for each video.

```
...
30 2 489 629 354 403 22
30 3 0 13 351 363 23
31 2 468 607 354 404 24
31 3 0 27 351 368 25
32 2 449 587 357 404 26
32 3 0 40 347 367 27
...
```

Figure 1: Example of the ground truth table

2.2 Evaluation Metrics

The next process (see Figure 2) is carried out as many times as estimated tracks a frame are returned by the tracking system. The information that is needed are the necessary data to evaluate the performance of our tracking system: time of prediction, track identifier, value of the minimum x and y and maximum x and y which surround the target (shape of a rectangle), the codes or identifiers of each ideal trajectories and a matrix to store the results. These data have been estimated by the tracking system explained in the former chapters.

First of all, the result tracks are checked to see if they match with the ground truth tracks registered in the ground truth table. For example, if the real image shows two aircrafts in the parallel taxiways while the tracking system displays three targets, the target which is in the middle of the screen (the 'ghost target'

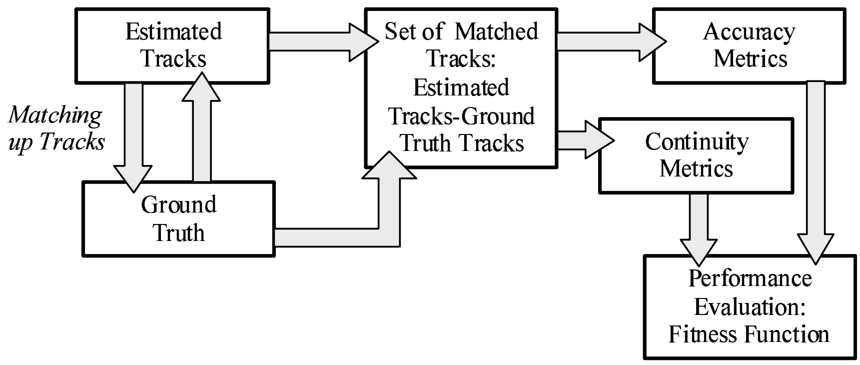

Figure 2: Calculation of Evaluation Metrics

mentioned in the former chapter) would not pass the test and it would be marked as a mismatched track.

Thus, if the test is passed, the next task of the evaluator is to find couples among all the tracks estimated for our system and the ground truth tracks. Once the couples are matched, the parameters to evaluate the quality of our system can be computed and stored in a matrix. The next list describes the quantities and how they are estimated. The parameters are divided in 'accuracy metrics' and 'continuity metrics'.

Accuracy Metrics:

1. *Error in area (in percentage)*:
   To begin with the first parameter, the difference between the ideal area and the estimated area is computed
2. *X-Error and Y-Error*:
   The difference among the x and y coordinates of the bounding box of an object estimated by the tracking system and the ground truth.
3. *Overlap between the real and the detected area of the rectangles (in percentage)*:
   The overlap region between the ideal and detected areas is computed and then compared, in percentage, with the original areas. The program takes the lowest value to assess the matching of tracking output and ground truth.

Continuity Metrics:

4. *Commutation*:
   One of the most important parameters to measure by the evaluator is the commutation. It is defined as follow: the first time the track is estimated, the tracking system marks it with an identifier. If this identifier changes in subsequent frames, the track is considered a commuted track. There could be several reasons why the identifier changes, but the most common is the loss of the track for a short time and subsequent recovery.

5. *Number of tracks*: It is checked if more than one detected track is matched with the same ideal track. If this happens, the program keeps the detected track which has a bigger overlapped area value, removes the other one and marks the frame with a flag that indicates the number of detected tracks associated to this ideal one.

3 Fuzzy Approach to Blob-to-Track Association Logic

When processing video output in dense airport areas, each available frame presents a set of blob-to-track multi-assignment problems to be solved, where several (or none) blobs may be assigned to the same track and simultaneously several tracks could overlap and share common blobs. So the association problem to solve is the decision of the most proper grouping of blobs and assignation to each track for each frame processed. Due to image irregularities, shadows, occlusions, etc., a first problem of imperfect image segmentation appears, resulting in multiple blobs potentially generated for a single target. So, blobs must be re-connected before track assignment and updating. This problem might be easily solved in single-target scenarios using a blob-grouping algorithm based on the blobs associated to the track in previous frames, defining a spatial gate for each track. However, when multiple targets move closely spaced, their image regions overlap, appearing some targets occluded by other targets or obstacles, so that some blobs can be shared by different tracks. So, a blob-to-track multi-assignment problem has to be solved, where several blobs could be assigned to the same track and simultaneously several tracks could overlap and share common blobs.

The traditional association systems use, together with motion estimation, target position (represented by centroids) extracted from sensor data. Conventional Nearest Neighbor systems [7] deals the assignment between plots and tracks as minimizing a global cost function. This function is computed based on the distance between plots and predicted tracks (residuals) and known statistical models for sensor errors. Bayesian extensions of NN, such as Multiple Hypothesis Tracking (MHT) [7] consider association decisions over several data scans, to ensure track continuity under critical conditions such as presence of false alarms, maneuvers or closely spaced targets. These types of hard-decision systems assume basic constraints of single plot updating each track, and no more than one track updated by the same plot, which are not applicable to the problem dealt.

A possible solution could be the removal of the one-to-one constraints and enumerate all possible grouping and assignment hypothesis, with approaches similar to that suggested in [12]. However, these types of solutions could demand excessive computation load to process in real time the frames and it would not ensure solving some problems such as the assignation of corrupted blobs resulting of the mix of several target images. As alternative, an all-neighbors approach, similar to Joint Probabilistic Data Association [7] or PMHT [15], seems adequate to this problem, since all blobs potentially gated with each track are used to update it, requiring besides quite lower memory and computation than MHT approaches. Other approaches apply the Expectation-Maximization [8] clustering algorithm for estimating the unknown correspondence among blobs and tracks. The groups of cells representing each target are modeled as a mixture of Gaussian pdfs of unknown parameters, so a likelihood function for those parameters given the measurements are computed at the same time as the unknown correspondence. The application of EM algorithm transforms the hard assignment in a continuous problem, numerically solved with a "hill-climbing" approach. It has been previously applied to data association for computer vision applications, and for a probabilistic approach to MHT, PMHT [15].

Traditional association systems represent targets with a single position and error parameters. Using a Video Surveillance System, an explicit representation of target shape and dimensions is more adequate to select the set of blobs gated by each track. Track-state vectors with position and cinematic estimates (2D location and velocity referred to the camera plane) are complemented with attributes defining a spatial representation of target extension and shape. So, the predicted target contour is used to gate blobs extracted in next frame.

There are not detailed models or analytical expressions to design this process, similar to Bayesian approaches for probabilistic association [9], but an analysis of continuity performance with different strategies, depending on numeric heuristics describing the situations, provide robust rules to take appropriate association decisions [12]. Rules have been obtained by analysis of performance under different conditions, characterized with these heuristics values. Rules represent the most proper actions to take under a set of particular extreme conditions to guarantee track continuity. Fuzzy reasoning techniques have been adopted to reproduce the system behavior under these conditions and besides generate the proper output for intermediate situations [2]. A fuzzy system [23] [16] is proposed to evaluate the confidence given to the information contained both in the gated blobs and predicted tracks, based on a set of numeric heuristics describing the characteristic of these multiple-blob-multiple-track association scenarios [13]. Besides, an automatic procedure (neuro-fuzzy technique, [17] [18]) was explored in order to extract rules directly from examples (expert decisions in extreme conditions) [14].

4 Evaluation of Fuzzy Systems in Real Conditions

The systems proposed with fuzzy association logic are compared with a crisp-decisions system [3] behaving as follows: it will update all blobs included in the gate if group density is higher than 0.7, otherwise, it will remove the farthest blobs from the group, and, if two or more tracks share any conflictive blobs, it will predict them without updating.

Regarding the fuzzy system, the input variables are [13]: (1) Overlapping heuristic, this component can be seen as a "soft gating", computed as the fraction of blob area contained within track predicted region; (2) group density and distance to track: this heuristic evaluates the ratio between areas of detected regions and non-detected areas (holes) in the finally reconnected pseudo-blob; (3) conflict with other tracks heuristic: this component evaluates the likelihood of blob being in conflict with other tracks (this problem appears when target trajectories are so close that track gates get overlapped and share the blob) and (4) proximity to image borders heuristic, finally, image borders are the areas where tracks are usually initialized, and so they are transient areas where tracks are not stabilized yet (this number evaluates if the blob is close to any of the four image borders). These heuristics provide useful information to be considered when assessing the confidence that may be given to each blob before track update. Additionally, the predicted track may be also characterized with some heuristics, indicating the confidence given to the fact that this track represents the motion of a real target, detecting when it is deviated from real trajectory. They are the following: (1) number of missed updates, it is the number of consecutive frames where no blob was included into track inner gate, (2) track detected area, conversely to blob overlapping heuristic, it is the proportion of area, within predicted inner gate, filled with blobs detected in current frame, and, (3) proximity to image borders, this value is equivalent to the one computed for blobs.

Although, these input variables are defined in the same way for the two systems, the membership function and the set of rules are defined in a different way. In the fuzzy system developed from the expert knowledge, the membership functions has trapezoidal shape and few rules are considered and, besides, rules have few combinations in the IF-part [13]. The learned fuzzy system [14] uses triangular shapes, with more rules than the previous one, and, with more sophisticated IF-part.

Next, we compare the evaluated performance for these four systems. The rigid scheme with crisp decisions is taken as a benchmark, and compared with the fuzzy systems, considering the three variants of rule sets mentioned above. This analysis has been performed on three representative scenarios described in [13], processed to obtain and store the reference ground truth. They are described next, together with the results.

4.1 Scenario 1. Conflicts and linear motion

In this scenario, there are several aircraft moving in parallel taxiways and their images overlap when they cross. This occurs always with uniform motion on straight segments. The results are presented in figure 3 for the four systems and a sample trajectory from an aircraft. The overlapped area and X-Y errors, with a direct relation (the higher the errors the lower the overlap between estimated area and ground truth), reflect tracking accuracy while tracking stability is in the number of real tracks representing the trajectory and commutations. The systems with the higher accuracy and stability here is the one with manual expert rules, and the refined one has a slight degradation. The rules obtained with neurofuzzy learning over examples achieve lower performance, and the hard-decision system has the worst accuracy.

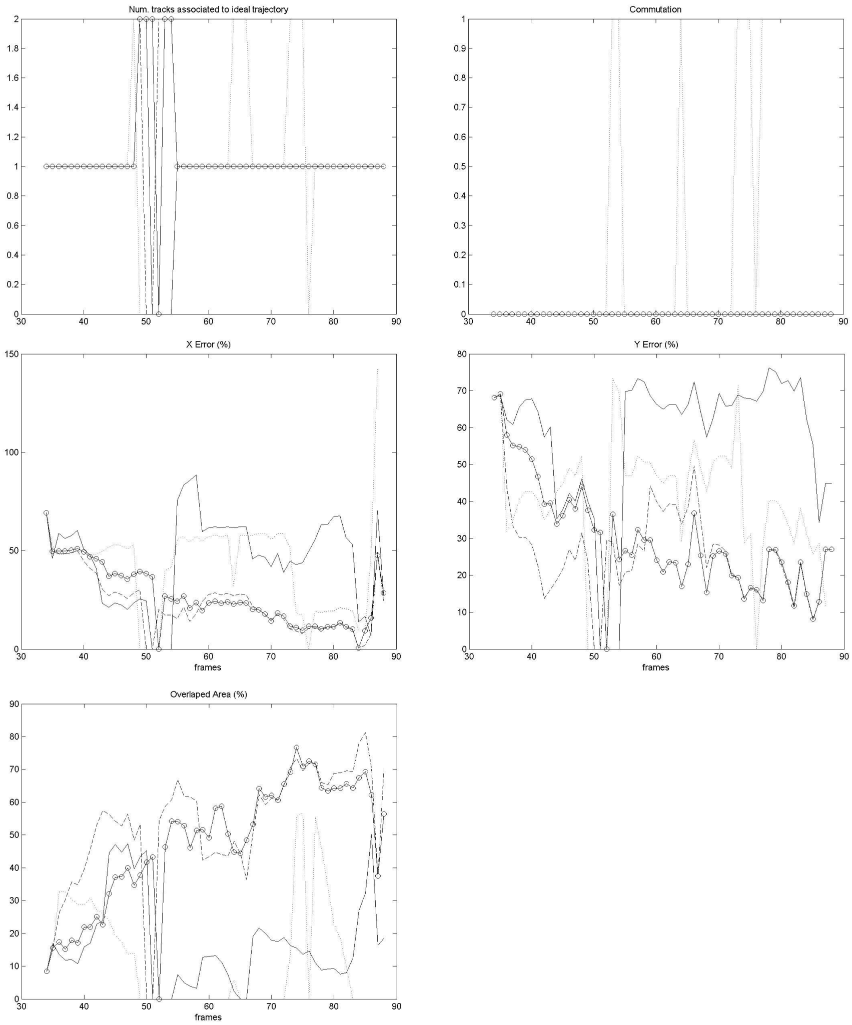

Figure 3: evaluation with scenario 1. - rigid scheme, – expert rules, neurofuzzy rules, o-o: tuned rules

4.2 Scenario 2. Conflicts with maneuvers

Two aircraft are moving on inner taxiways between airport parking positions and one of them is occluded during 25 frames. Besides, both aircraft are turning during the conflict interval, changing their orientations. The presented results are for the occluded aircraft. In this case, the tuned rules achieve better accuracy than the previous system. The one with neurofuzzy learning has segments with better accuracy, but its stability is clearly worse. The worst performance is again for the crisp-decisions system.

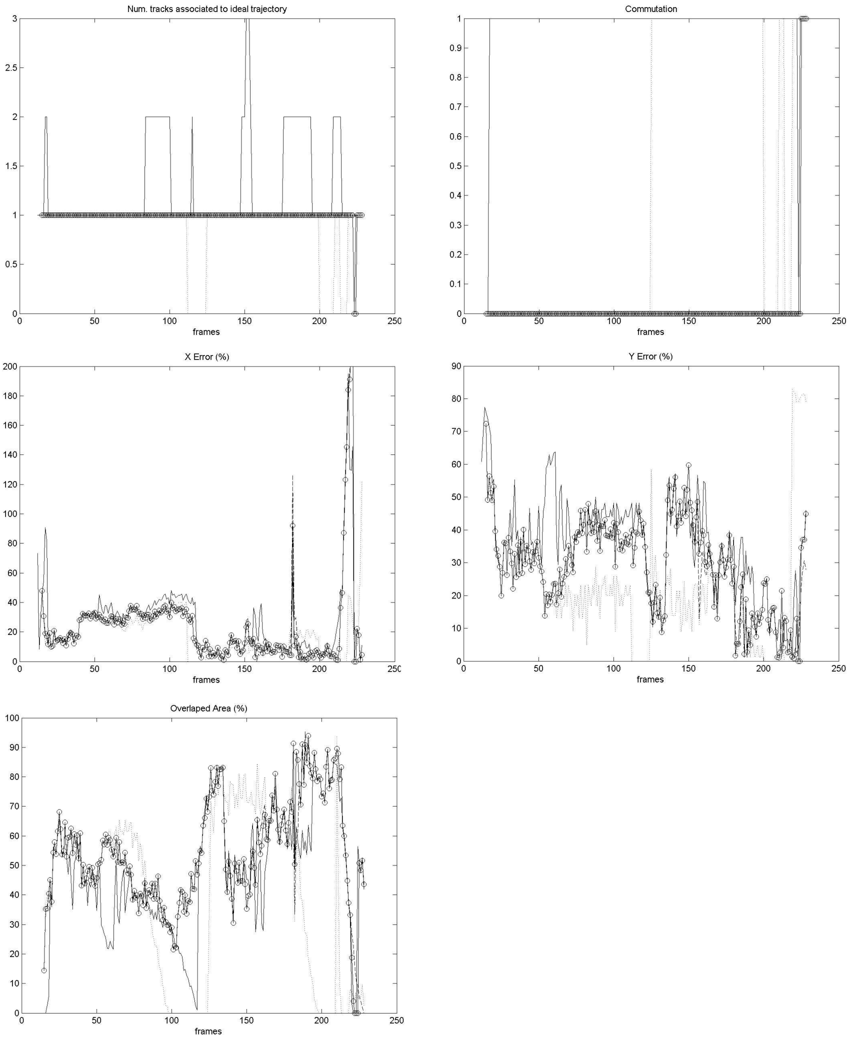

Figure 4: evaluation with scenario 2. - rigid scheme, – expert rules, neurofuzzy rules, o-o: tuned rules

4.3 Scenario 3. Occlusions and target fragmentation

This is a multiple-blob reconnection scenario. There is an aircraft moving with partial occlusions due to stopped vehicles and aircraft in parking positions in front of the moving object. There are multiple blobs representing a single target that must be re-connected, and at the same time there are vehicles in parallel roads that must be kept separated from this trajectory. In this scenario, the tuned rules achieve the best accuracy and stability.

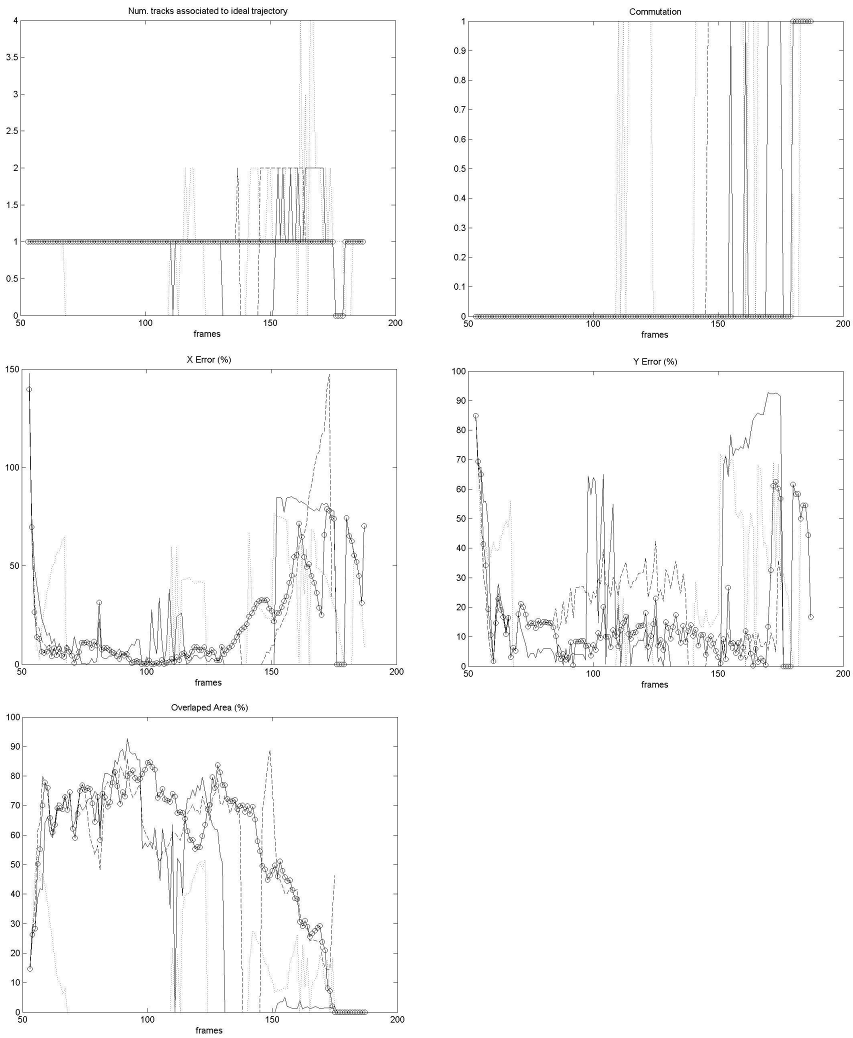

Figure 5: evaluation with scenario 3. - rigid scheme, – expert rules, neurofuzzy rules, o-o: tuned rules

5 Conclusions

We have presented several fuzzy approaches to solve the blob-to-track association problem. In this work, a novel process to evaluate the performance of a tracking system based on the ground truth extracted of information from images filmed by a camera has been developed. The ground truth tracks, which have been previously selected and stored by a human operator, are compared to the estimated tracks. The comparison is carried out by means of a set of evaluation metrics which are used to compute a number that represents the quality of the system. This process allows the comparison of different association logics using hard decision or fuzzy decisions. The main problems to solve with real-world examples are identified and a further refinement of previous rules was possible to improve the output stability. A further work will explore the application of optimization strategies making use of this method to improve the design.

6 Acknowledgements

The research reported in this paper is partly supported by CICYT (TIC2002-04491-C02-02) project.

References

[1] *European Manual of Advanced Surface Movement and Control Systems (AS-MGCS)*, Volume 1: Operational Requirements. Draft Version 04, August 2000.

[2] H.R. Berenji and P. Khedkar, *Learning and Tuning Fuzzy Logic Controllers through Reinforcements*, IEEE Transactions on Neural Networks **3**, 724-740, 1992.

[3] J. A. Besada, J. Portillo, J. García and J. M. Molina, *Image-Based Automatic Surveillance for Airport Surface* in procs. FUSION 2001 Conference. Montreal, Canada. August 2001.

[4] J. A. Besada, J. Portillo, J. García, J. M. Molina, A. Varona and G. Gonzalez, *MARIA: Preliminary results of an AENA R&D project on airport surface surveillance based on CCTV* in procs. International Conference on Airport Surveillance Sensors. December 2001.

[5] J. A. Besada, J. M. Molina, J. García, A. Berlanga and J. Portillo, *Aircraft Identification integrated in an Airport Surface Surveillance Video System.* Machine Vision & Applications **15** (3). July 2004.

[6] J. Black, T. Ellis and P. Rosin. *A Novel Method for Video Tracking Performance Evaluation*, in procs. Joint IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS), 2003.

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

[8] A. Dempster, N. Laird and D. Rubin, *Maximum likelihood from incomplete data via the EM algorithm.* Journal of the Royal Statistical Society **39** (Series B), 1-38, 1977.

[9] Z. Ding, H.Leung and L. Hong, *Decoupling joint probabilistic data association algorithm for multiple target tracking* IEEE Radar, Sonar and Navigation **146** (5), 251-254, October 1999.

[10] *Proceedings of the ECAC APATSI and EC Workshop on A-SMGCS* Frankfurt, Germany. April 1994.

[11] U.S. Department of Transportation FAA, *The Future Airport Surface Movement Safety, Guidance and Control Systems: A vision for Transition into the 21st Century*, Washington, U.S., November 1993.

[12] J. García, J. A. Besada, J. M. Molina and J. Portillo, *Fuzzy data association for image-based tracking in dense scenarios*, in procs. IEEE International Conference on Fuzzy Systems. Honolulu, U.S., May 2002.

[13] J. García, J. M. Molina, J. A. Besada and J. Portillo, *Fuzzy Approach for Data Association in Image Tracking*, Mathware and SoftComputing. **X** (2-3) pp 117-129, 2003.

[14] J. García, J. M. Molina, J. A. Besada and J. Portillo, *A MultiTarget Tracking Video System based on Fuzzy and Neuro-fuzzy Techniques*, EURASIP Journal on Applied Signal Processing, special issue on Advances in Intelligent Vision Systems: Methods and Applications (in press).

[15] H. Gauvrit, J.P le Cadre and C. Jauffret, *A Formulation of Multitarget Tracking as an Incomplete Data Problem*, IEEE Transactions on Aerospace and Electronic Systems. October 1997.

[16] J. M. Mendel, *Fuzzy Logic Systems for Engineering: A Tutorial*, in procs. IEEE. **83** (3),. pp 345-377, March 1995.

[17] D. Nauck and R. Kruse, *NEFCLASS- a neuro-fuzzy approach for the Clasification of Data*, in procs. ACM Symposium on Applied Computing, pp 461-465, Nashville, U.S., 1995.

[18] D. Nauck, R. Kruse and R. Stellmach, *New Learning Algorithms for the Neuro-Fuzzy Environment NEFCON-1*, in procs. Neuro-Fuzzy Systems'95, 357-364, Armstadt, Germany, 1995.

[19] J. H. Piater and J. L. Crowley, *Multi-Modal Tracking of Interacting Targets Using Gaussian Approximations*, in procs. IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS), 2001

[20] D. Pokrajac and L. J. Latecki, *Spatiotemporal Blocks-Based Moving Objects Identification and Tracking*, in procs. IEEE Int. Workshop Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS), October 2003.

[21] P.L. Rosin and E. Ioannidis, *Evaluation of global image thresholding for change detection*, Pattern Recognition Letters **24** (14), pp. 2345-2356, 2003.

[22] C. E. Schwabn and D. P. Rost, *Airport Surface Detection Equipment*, Proceedings of the IEEE **2**, February 1985.

[23] L. A. Zadeh, *Outline of a New Approach to the Analysis of Complex Systems and Decision Processes*, Transactions on SMC **3** (1), January 1973.