A PROTOTYPE FUZZY SYSTEM FOR SURVEILLANCE PICTURE UNDERSTANDING

HELMAN STERN, URI KARTOUN*, ARMIN SHMILOVICI

Department of Industrial Engineering and Management, Ben-Gurion University
P.O.Box 56, Be'er-Sheeva 84105, ISRAEL
Fax: +972-8-6472958; Tel: +972-8-6461434
E-mail: (helman, kartoun, armin)@bgumail.bgu.ac.il

ABSTRACT

The last stage of any type of automatic surveillance system is the interpretation of the acquired information from its sensors. This work focuses on the interpretation of motion pictures taken from a surveillance camera, i.e.; image understanding. A prototype of a fuzzy expert system is presented which can describe in a natural language like manner, simple human activity in the field of view of a surveillance camera. The system is comprised of three components: a pre-processing module for image segmentation and feature extraction, an object identification fuzzy expert system (static model), and an action identification fuzzy expert system (dynamic temporal model). The system was tested on a video segment of a pedestrian passageway taken by a surveillance camera.

Keywords: *image understanding, picture segmentation, fuzzy expert systems, surveillance video*

1. INTRODUCTION

With the continuous decline in the price of imaging technology, there is a surge in the use of automatic surveillance systems and closed circuit TV (CCTV). Banks, ATM machines, schools, hospitals, transport walkways employ automatic video recording of their surrounding environments. There appears to be little human inspection (in real-time or otherwise) of these surveillance videos, and thus the system is relegated to a simple deterrence function (mainly for deterrence of possible felonies). However, in many environments it is necessary to understand the contents of the video for subsequent event detection, storage and retrieval. Extraction of the desired events requires a high semantic level of human understanding and requires a prohibitive amount of human processing.

Automatic processing of surveillance videos introduces several practical problems. Recording, storing and managing of large volumes of data is expensive, and cost effective solutions are not yet available for cases where most of the data is useless. Also, in the case that someone will want to inspect the contents of the video, there would be a great deal of work involved in watching all the recorded segments. Thus, there is a need for an effective means by which the content of the data can be automatically characterized, organized, indexed, and retrieved, doing away with the slow, labor-intensive manual search task. Understanding the contents of an image, in the context of it's importance to the operator of the surveillance system, is the key to efficient storage and retrieval of video segments.

The problem of automatic image understanding is a difficult one. There are two possible paradigms for this problem [1,2]: computational feature based semantic analysis - the detection of features based on elaborate computational models; and human cognitive perception of high level semantics - the subjective user interpretation of features in the image.

With the computational paradigm, it is technically difficult to identify correctly and in a reasonable amount of time, the contents of an image in all possible circumstances (*e.g.*, identify an object from all possible angles). There is a need to develop a model for the features of an image in all possible circumstances.

Human cognitive perception, on the other hand, starts with simple object segmentation, that is to segment 2D-plane images into physically meaningful objects. Image understanding is related to the relations between the objects in the picture, and the context in which they appear. This, in general, is very difficult to formulate as a computational problem. Yet, people, even children, can learn to do it with ease. The problem of image understanding can be facilitated [3,4] if we can restrict the type of objects to be identified, (*e.g.*, humans), the quality of the identification (*e.g.*, contours only), the possible relations between objects (*e.g.*, approaching each other), and the context in which they operate (*e.g.*, a closed passageway).

In this work a prototype of fuzzy expert system is presented which can describe, in natural language like-way, simple human activity in the field of view of a

surveillance camera. The system has three different components: a pre-processing module for image segmentation and feature extraction, an object identification fuzzy expert system (static model), and an action identification fuzzy expert system (dynamic temporal model). The system was tested on a video segment of a pedestrian passageway taken by a surveillance camera

The rest of this paper is organized as follows: Section 2 describes the surveillance problem and its simplification. Section 3 describes the construction of the fuzzy expert system. Section 4 explains the static and dynamic fuzzy expert systems for object identification and object behavior, respectively. Section 5 provides the results of applying the system to a pedestrian passageway. Section 6 concludes the paper with some discussion.

2. PROBLEM DEFINITION

Figure 1 presents a typical scene that might be observed under a surveillance system. This scene is a semi-covered passageway between two buildings at Ben-Gurion University. The open door on the lower left leads to a cafeteria, while there is a lecture hall on the right side. Sunlight enters through the right side of the scene during daytime, so the lighting conditions can change drastically during the day causing variable length shadows to be present in this scene. It was decided to use a single camera whose visual axis points down into the scene from an observation point. This affords the use of prior knowledge regarding a minimalist modeling, for future real-time processing that exploits complimentary qualities of different visual clues, such as relative size and motion of objects.

Figure 1. A typical scene taken from a surveillance camera

In the above scene, people are the objects of interest, though occasionally other objects appear such as, an electric delivery cart to the cafeteria and birds. It is desired to describe the activities in this scene in terms of the activities of the people.

The full range of possible human activities that can be described by a natural language is very large. Fortunately, in the context of surveillance, there is interest only in a small subset of these human activities. This project focuses on the identification of normal human behavior. In a future project, we will address the abnormal behavior aspects such as; trespassing, violence, and theft. In the context of the scene above, people can be either standing, walking, or running. They can be found alone or in groups of two or more. In either case, the position of the person or group is described relative to the scene.

As it turns out, it is still technically difficult to identify the concept of a “person” in a noisy environment such as that described above, especially with artifacts due to changing lighting conditions, shadows, optical distortions, reflections on large glass windows, and the variability of pedestrian motions. Also, there is a limit on the computational time needed to generate a description. Thus, the following simplifications to the problem were made [5,6]:

(i) A primitive notion of a “blob” is defined as a set of clustered pixels that have moved between two given images. The blobs in each image were segmented from the background of the image. A blob may or may not be a real object in the scene such as a person. It may instead be background noise that was not removed during the image segmentation operations. This simplification, however, facilitated the necessary image pre-processing operations.

(ii) A fuzzy inferencing mechanism is used for perceptual integration of simple visual clues. The theory of fuzzy sets is employed which can use a linguistic variable such as distance, for example. This linguistic variable can take on various terms such as; very close, close, far, etc. This replaces the “crisp” mathematical description of distance such as, for example, 4.15 meters. This facilitated the use of mathematical models that capture and describe the activities in the scene in terms of natural like language.

The goal of this project is defined as follows: develop a prototype of a fuzzy expert system that can understand and describe a scene in natural like language. Fuzzy rules are based on domain expert knowledge to describe a scene, locations of objects of interest, object descriptions and object behaviors.

Given a set of scenes (images) from a video clip taken at consecutive times, describe the scene in terms of number of people, and people groups in the scene and their actions such as; walking toward or away from the camera, standing still, departing from another person, walking with another person, joining a group, etc.

3. FUZZY EXPERT SYSTEMS

Fuzzy set theory [7] and fuzzy expert systems [8] are used to capture expert knowledge that cannot be easily formulated mathematically, or when the

mathematics is too complex to solve. Building an expert system starts with interrogating domain experts (in this case, image processing experts and surveillance staff) and formulating their knowledge in the form of linguistic variables and fuzzy rules [7]. Additional domain knowledge can be included from other sources. Also, some knowledge can be derived from statistical analysis of historical information. References [9-11] present applications of fuzzy logic to image processing.

The fuzzy system considered in this paper is comprised of four basic elements [7]: a fuzzifier, a fuzzy rule base, a fuzzy inference engine, and a defuzzifier. We consider multi-input single-output fuzzy systems as elaborate mapping functions: $f: U \subset \boldsymbol{R}^n \rightarrow V \subset \boldsymbol{R}$, where $U = U_1 \times U_2 \times \ldots \times U_n \subset R^n$ is the input space and $V \subset \mathbf{R}$ is the output space. A multi-output system can be represented as a group of single-output systems.

A rule is a proposition that implies another proposition. In this paper, the *fuzzy rule base* consists of a set of linguistic rules in the form of "IF a set of conditions is satisfied THEN a set of consequences is inferred". Assume that there are *N* rules of the following form:

R_i: IF x_1 is A_{i1} and x_2 is A_{i2} and...and x_n is A_{in} THEN y is C_i, $i=1,2,...,N$

where x_j $(j=1,2,...,n)$ are the input variables to the fuzzy system, y is the output variable of the fuzzy system, and the fuzzy sets A_{ij} in U_j and C_j are linguistic terms characterized by fuzzy membership functions $A_{ij}(x_j)$ and $C_i(y)$, respectively. Each rule R_i can be viewed as a *fuzzy implication* (relation) $A_i = A_{i1} x A_{i2} x ... x A_{in} \rightarrow C_i$, which is a fuzzy set in $U \times V = U_1 \times U_2 \times ... \times U_n \times V$ with membership function $R_i(\bar{x}, y) = A_{i1}(x_1) * A_{i2}(x_2) * ... * A_{in}(x_n) * C_i(y)$, and* is the T norm [7], $\bar{x} = (x_1, x_2, ..., x_n) \in U$ and y $\in$ V.

4. THE STATIC AND DYNAMIC FUZZY EXPERT SYSTEM MODELS

In this section the static and dynamic (temporal) expert systems are described. Initially, however, a brief discussion of the pre-processing stage is given although this is not the main focus of the paper. In the *Pre-processing* stage, a raw image is pre-processed for the identification of blobs. The main function of the static fuzzy expert system is for *object identification*. It uses the geometrical attributes of the blobs to make inferences about the objects in the picture. The main function of the dynamic fuzzy expert system is for *action identification.* It uses temporal movement attributes of objects to make inferences about the behaviors of the objects in the picture.

4.1 Image Preprocessing

The pre-processing stage starts with a grabbed gray-scale image of a scene. Various image processing operations are used; removal of noise from the image due to optical distortions of the lens, adaption to ambient and external lighting conditions, and segmentation of the blobs from the background. The end result is an image of segmented blobs from which features are extracted. Using the Image Processing Toolbox of MATLAB, eleven different geometrical attributes were defined for each blob in the image:

Area - the actual number of pixels in a blob. *Convex area* - the number of pixels in the convex area of a blob. *Solidity* - the ratio between the above two area measures. The *Equivalent Diameter* of a circle with same area. *Centroid* - the coordinates of the center of gravity of the blob. The coordinates of the *Bounding Box*. The *Minor Axis Length*, *Major Axis Length*, and *Eccentricity* for a bounding ellipsoid. The *Orientation* (in degrees) to the horizontal axis. *Extent* - the proportion of the pixels in the bounding box that are also in the detected blob. Figure 2 presents a segmentation of the scene shown in figure 1.

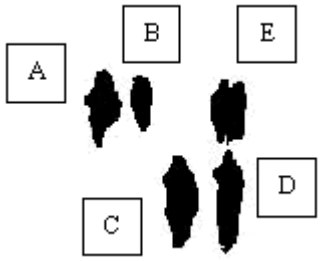

Figure 2. Segmentation of figure 1 into blobs

Different features are associated with each blob. *Static features*, such as blob centroid and size, will be used to classify each blob in the scene into separate categories such as: *one person*; *two people*; *more than two people* or *a noise blob*. These are discussed further in the section on the static model. *Dynamic features*, such as *direction of movement* relative to the camera, will be used to classify the activities of each blob in the scene into categories such as: *blob moving toward* camera, *blob moving away* from camera, blob is stationary. Other possible descriptions, not investigated here, result from the relational positions between different blobs, such as: blob has merged with another blob; a blob has split into two blobs; a close pair of blobs walk together; two blobs meet and stop; two blobs depart from each other and move in different directions, etc. These are discussed further in the section on the dynamic model.

4.2 The Static Fuzzy Expert System Model

For the object identification stage, a fuzzy expert system was built, with three input variables and one

output (conclusion) variable. Linguistic variables were defined for each input.

- The Area (in pixels) of the blob, defined on the range [0, 3000] pixels, can take on five terms: Area = {*very-small, small, normal, large, very-large*}
- The aspect ratio of the bounding box (the height/width ratio of the smallest rectangle that can contain a blob), defined on the range [0, 3.5], can take on five terms: Ratio = {*very-small, small, normal, large, very-large*}
- The y-coordinates of the center of mass of a blob (a simple estimator for the distance from the camera), defined on the range [0, 250] pixels, can take on four terms: Distance = {*very-far, far, close, very-close*}
- The conclusion about the identity of the blob, defined on the range [0, 30], can take on five terms: Conclusion = {*not-a-person, single, couple, three, many*}

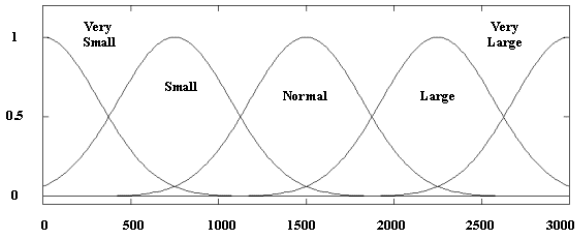

Figure 3(a). Membership input function for size of a blob

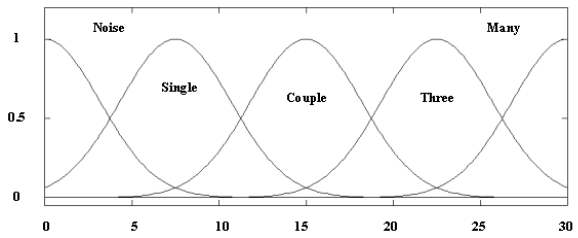

Figure 3(b). Membership output function for blob type

Figures 3(a) and 3(b) display the fuzzy sets defined for two of the linguistic variables. Those were selected as Gaussian membership function from Matlab's fuzzy logic toolbox [13] after some experimentation. Table 1 presents part of the rule-base for blob type identification type. These seemingly simple rules are logically derived from prior knowledge of the relative position of the camera and the scene. Also, the rules utilize implicit relational knowledge about the image such as; "a far object has small area", or "groups have larger area than a single person". While it is possible to formulate mathematical functions for these notions, they most likely will be complex because of the stochastic nature of the relationships.

4.3 The Dynamic Fuzzy Expert System Model

For the action identification stage, a second fuzzy expert system is defined with two input variables and two output (conclusion) variables. Linguistic variables are defined to represent the temporal aspects of the blobs.

- The X-movement change - the change of the centroid of a blob in the x-axis of the camera image plane. Defined on the range [-5, +5] pixels, can take on five terms: X-movement = {*dramatically-left, slightly-left, almost-no-change; slightly-right; dramatically-right*}.
- The Y-movement change - the change of the centroid of a blob in the y-axis of the camera image plane, defined on the range [-5, +5] pixels, can take on four terms: Y-movement = {*dramatically-away, slightly-away, almost-no-change; slightly-forward, dramatically-forward*}.
- The conclusion about the object's speed, defined on the range [-2, +2] units can take on four terms: Speed = {*standing, slow, fast, running*}.
- The conclusion about the object's direction, defined on the range [0,1] can take on eight terms: Direction = {*right, forward, backward-left, backward-right, forward-right, forward-left, backward, left*}.

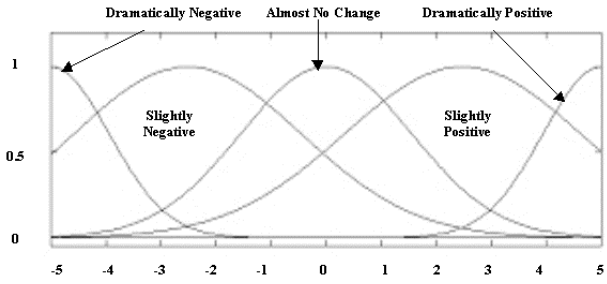

Figure 4. Membership functions for the x-axis change

Figure 4 presents an example of the membership functions defined for one of the linguistic variables called x-axis change. Table 2 presents part of the rule base used for blob action identification.

The fuzzy system was implemented with MATLAB's "fuzzy logic tool box". The operation of the fuzzy system used in this paper was restricted to "Centroid defuzzifier" - that is the center of gravity operator was used to combine the different values resulting from the activation of each fuzzy rule into one crisp result.

Note, that this is a prototype system only, so some choices (*e.g.*, the shape of the membership functions) are rather arbitrary. Though further optimization is possible, one of the typical characteristics of fuzzy expert system is that their predictions are reasonable even with a minimal number of assumptions, and a partial rule-base.

5. APPLICATION TO UNDERSTANDING HUMAN ACTIVITIES IN A PEDESTRIAN PASSAGEWAY

A video clip of approximately 12 seconds (25 frames per second) was taken from the surveillance camera located at Ben-Gurion University. The video clip was divided into 299 frames. The pre-processing stage was performed using Matlab's Image Processing Toolbox. This stage generated all the blobs and their features.

Several sets of frames were used to test the performance of the fuzzy systems for different scenarios. In this paper, Frames 170 (figure 1) and 180 (figure 5) are used for demonstrating the behavior of the action identification dynamic fuzzy expert system described in Table 2. Figure 6 is the segmentation of frame 180 (shown in Figure 5) of the video.

Figure 5. Frame no. "180"

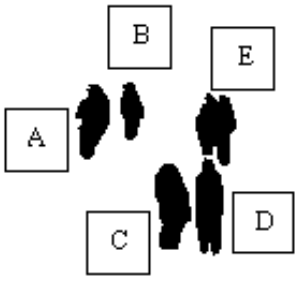

Figure 6. The segmentation of frame 180

Table 3 presents the values of the geometrical attributes; *Area*, *BoundingBox* and *Centroid* for the blobs of frames 170, and 180. The conclusions of the static expert system (of Table 1) regarding the identity of the blobs in figure 2 are presented in bold letters. For example, blob B is correctly identified as a *single* person. Blob E is correctly identified as a *couple*.

The last three rows in Table 3 present the input and output variables from the dynamic fuzzy expert system. Based on tracking the outputs of the static expert system the attributes of *X,Y movements* are determined. The speed and the direction of each blob are identified and shown in the last two rows of the table. For example, blob A was identified as *standing*. Blob C was identified as moving slowly in the *forward-right* direction. Thus, the content of the scene could be summarized with five natural like language statements such as; "A *single* person in location (x,y) is moving slowly in the *forward-right* direction". Furthermore, tracing blobs over time, can be used to identify splitting and merging events such as; "one person + two people = three people".

6. CONCLUSIONS AND DISCUSSION

A fuzzy expert system has been developed for high-level understanding of images from a surveillance video. Two fuzzy inference systems were described - static and dynamic. In the implementation of these two systems the static model is used to identify blobs as clusters of people and provide an estimate of the number of people in each blob. The dynamic model uses temporal information between frames to obtain movement information of the blobs, and with little additional effort can place existing blobs into new categories by identifying merge and split activities. Although, correct blob identifications were made for the frames examined; further testing is required. The evaluation of such extended testing requires a performance measure such as; percent of corrected blob types (for the static model), and mean square velocity error (for the dynamic model).

Although not implemented here, it is possible to display a "degree of belief" of the expert system regarding the validity of its conclusions. The degree of belief is the value of the membership of the output variables. The larger the membership value, the higher the level of belief. Low degrees of belief might encourage the collection of additional information before presenting the conclusion. Also, further optimization of the fuzzy system will reduce the probability of false classifications.

The information may be placed in a database to record and update a list of people in the scene. People will be dynamically added and deleted from the list. This can act as a "pedestrian log". The database can then be used to index the images for archival purposes and subsequent retrieval. For example, each blob in Figure 5 will be indexed by its identified class, and its identified actions. It is possible to store only the image information in the bounding box of each blob, for image compression, with minimal information loss. Other expert systems can use this image information to make inferences about other types of activities such as violence, vandalism or theft. It may also be used to study behavioral aspects of crowds or pedestrians. For example, one can study "who gives way to whom", and "do large blobs act as attractors for passerby's"?

References

[1] Hanjalic, A. and Li-Qun Xu, Video Analysis and Retrieval at Affective Level, submitted to Special Issue of *IEEE Transactions on Multimedia Database*, April 2001.

[2] Xu, Li-Qun, Jian Zhu and Fred Stentiford, Video Summarization and Semantics Editing Tools, in *Proceedings of SPIE on Storage and Retrieval for Media Databases 2001*, San Jose, CA, January, 2001.
[3] ICONS: Incident Recognition for Surveillance and Security, DTI/EPSRC LINK Project, July 2000-2003, *http://www.dcs.qmw.ac.uk/research/vision/projects/ICONS/*
[4] VIGOUR - An Integrated Vision System for Research in Human Recognition, The ISCANIT Project, *http://www.dcs.qmw.ac.uk/research/ISCANIT*.
[5] S. J. Dickinson, H. I. Christensen, J. K. Tsotsos and G. Olofson, Active Object Recognition Integrating Attention and Viewpoint Control. *Computer Vision and Image Understanding*. 67(3), 1997, pp. 239-260.
[6] R. Lengagne, T Jean-Philippe. and M. Olivier, From 2D Images to 3D Face Geometry, *Proceedings of IEEE Second International Conference on Automatic Face and Gesture Recognition*, 1996.
[7] H. Zimmerman, "*Fuzzy Set Theory*", Kluwer, 1991.
[8] E. Cox, "*The Fuzzy Systems Handbook*", (Boston, AP professionals, 1994)
[9] J. Shanahan, B.Thomas, M. Mirmehdi, N. Campbell, T. Martin and J. Baldwin, Road Recognition Using Fuzzy Classifiers, Advanced Computing Research Center University of Bristol, 2000.
[10] M Balsi. and F. Voci, Fuzzy Reasoning for the Design of CNN - Based Image Processing Systems. *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS 2000)*, Geneva, Switzerland, 2000.
[11] P.Matsakis, J. Keller, L. Wendling, J. Marjamaa and O. Sjahputera, Linguistic Description of Relative Positions in Images. University of Missouri-Columbia, Columbia, USA., 2000.
[12] Matlab software, Image Processing Toolbox User's Guide. *The MathWorks, Inc.* (http://www.mathworks.com), 1998.
[13] Matlab software, The Fuzzy Logic Toolbox User's Guide, *The MathWorks Inc*., (http://www.mathworks.com), 2000.

Premise part	Conclusion part
If Area = *very-small*	Then *not-a-person*
If Distance = *close* and Area = *small* and Ratio = *normal*	Then *single-person*
If Distance = *close* and Area = *small* and Ratio = *very-large*	Then *single-person*
If Distance = *close* and Area = *normal* and Ratio = *large*	Then *single-person*
If Distance = *close* and Area = *normal* and Ratio = *very-large*	Then *single-person*
If Distance = *close* and Area = *large* and Ratio = *normal*	Then *single-person*
If Distance = *far* and Area = *very-small* and Ratio = *very-large*	Then *single-person*
If Distance = *far* and Area = *small* and Ratio = *very-large*	Then *a-couple*
If Distance = *far* and Area = *normal* and Ratio = *large*	Then *three-people*

Table 1. Fuzzy rules for blob type identification

Premise part	Conclusion part
If X-movement = *slightly-right* and Y-movement = *almost-no-change*	Then Speed = *standing* and Direction = *none*
If X-movement = *slightly-right* and Y-movement = *slightly-forward*	Then Speed = *slow* and Direction = *forward-left*
If X-movement = *slightly-right* and Y-movement = *dramatically-forward*	Then Speed = *fast* and Direction = *forward-leftt*
If X-movement = *almost-no-change* and Y-movement = *almost-no-change*	Then Speed = *standing* and Direction = *none*
If X-movement = *dramatically-left* and Y-movement = *almost-no-change*	Then Speed = *fast* and Direction = *right*

Table 2. Fuzzy rules for blob action identification

Blob Identification	**A**	**B**	**C**	**D**	**E**
Area (170) (fig. 1)	378	195	490	401	458
BoundingBox (170) (fig. 1)	[128.5 36.5 17 37]	[149.5 39.5 10 26]	[164.5 77.5 17 43]	[186.5 41.5 17 31]	[187.5 74.5 15 49]
Centroid (170) (fig. 1)	[136.6 53.9]	[154.7 51.7]	[173.1 98.1]	[195.2 56.5]	[195.1 97.9]
Fuzzy Blob Type Conclusion	**single**	**single**	**single**	**single**	**couple**
Area (180) (fig. 5)	369	204	489	434	474
BoundingBox(180) (fig. 5)	[129.5 37.5 16 35]	[150.5 36.5 11 27]	[166.5 74.5 17 40]	[185.5 41.5 19 34]	[185.5 71.5 13 46]
Centroid (180) (fig. 5)	[137.7 54.1]	[155.6 49.7]	[174.4 94.1]	[195.1 57.5]	[191.9 95.0]
Fuzzy Blob Type Conclusion	**single**	**single**	**single**	**single**	**couple**
X,Y-movements	1.03 0.19	0.92 -1.99	1.35 -3.99	-0.04 1.03	3.17 -2.95
Fuzzy Blob Speed Conclusion	**stand**	**stand**	**slow**	**stand**	**slow**
Fuzzy Blob Direction Conclusion	**None**	**None**	**forward-right**	**None**	**forward-right**

Table 3. Outputs of fuzzy systems regarding blobs in figures 2, 6