An ontology-based approach to human telepresence

Vincenzo Di Lecce*, Marco Calabrese
DIASS
Politecnico di Bari
Taranto, Italy
{v.dilecce, m.calabrese}@aeflab.net

Vincenzo Piuri
DTI
Università degli Studi di Milano
Crema, Italy
piuri@dti.unimi.it

***Abstract*—Detecting human presence automatically is a challenging task since several environmental parameters may affect the quality and the continuity of detection. Although many techniques have been developed so far in the literature to solve this problem, they generally rely on well-defined operational context. Hence, they are sensitive to uncontrolled variables and unpredicted events. In this work an ontology-based approach to human telepresence detection is presented. Contrarily to classic sensor-driven techniques, a top-down methodology is applied. Starting from a formal description of the problem ontology, a set of high-response rate and low-response rate sensors is employed in a computational model. As a consequence of this model, a multi-sensor equipped device has been experimentally setup to conduct measurements on real scenarios. Experiments have been devised to estimate the robustness of the detection. In particular, some preliminary evaluations related to using a minimal set of chemical sensors are reported.**

Keywords- human telepresence detection; continuity of detection, ontology-based approach

I. INTRODUCTION

There can be found several situations where the detection of the human presence is of certain relevance. From hazardous material handling in manufacturing workplaces to indoor pc workstations, a system being able to answer the simple question: "is there a human out there?" can be useful. A correct response requires however different levels of abstraction. Nevertheless, the inherent limitation provided by the semantic gap makes human detection a difficult task to automate. The automation of human detection is, in fact, sensitive to several environmental parameters such as rapid movements of the target or cluttered backgrounds which affect the continuity of detection.

The same problems can be found in telepresence environments. In [15] Hill and Jensen define telepresence as "an enhanced form of teleoperation that employs an immersive and transparent user interface, permitting the user to work with high effectiveness in inaccessible or remote environments". In this case, a system that judges, without human intervention, if a human being is present or not in a certain area, is appealing for many practical situations.

The simplest solution to human detection is represented by human operators directly monitoring surveilled areas. This is at the opposite side of detection automation: reporting the apparent drawback of such an approach is unworthy of mention. Generally, in the literature the problem of human detection is faced starting from a preliminary categorization between indoor and outdoor environment. In both cases, the approach that seems to prevail is that of trying to find physical signals that make the human presence evident from temporal sequence analysis. Automatic image-based processing techniques are countless in the literature to solve this problem, but they generally rely on well-defined operation contexts. Their enumeration will be skipped for the sake of conciseness. Nishi et al. [1], use radio transmission wave receivers in UHF-TV and VHF-FM bands to detect human presence or absence because of the signal fluctuation resulted from multi-path fading or shadowing with human presence. Analogously, in [2] the authors exploit the emissivity of the human body in the millimeter-wave band in contrast to more traditional infrared measurement techniques, which are quite sensitive to the time of the day (maximal thermal differences between a living body and the background are spotted during nighttime). Although quite intriguing from the perspective of pattern analysis, these approaches suffer profoundly from the semantic gap. There can be, in fact, other objects different from the human being having similar absorption coefficients; it is difficult (if not impossible) to discriminate the exact number of people in crowded areas; it is hard to infer on which kind of movement the detected person is doing and so on. To state if a person is present or not in a certain environment an external reference knowledge is then required. For this reason, in [6] a digital camera is employed. From a formal logic point of view, it is possible to state that approaches of this kind have a *propositional* nature (on-off logic) but fail to provide a *predicative* description.

Another prevailing idea in the literature is to focus on context-sensitive issues. In [3] for example human detection techniques in non-urban areas are discussed in contrast to urban ones. In [4] the additional effect of mounting detector to a moving platform is considered. Moreover, some scenarios as for video surveillance systems require additional features like person tracking [5]. Additionally, whenever two or more information sources come into the play some kind of data fusion technique is needed [9, 13, 14]. In all these cases of course the proposed methods are not algorithmic solutions to the general problem of human detection in all its aspects. Nonetheless, they represent empirical techniques that adopt a case-specific procedural approach depending on the parameters under investigation.

This work settles in the line of ontology engineering as a top-down approach [7]. Starting from a preliminary description

* corresponding author

of the problem ontology at an abstract level, the minimal set of used sensors as a consequence of such description (rather than an independent input for the measurement task) is derived and employed in a computational model. A set of tests is successively setup in order to statistically evaluate the robustness of the provided model against real-world scenarios. In this way, numerical pieces of information coming from multiple sources are treated as different symbols of the same given language-based representation. Local operation is directed to support privacy protection for the users' biometries.

The outline of the paper is as follows: Section II illustrates the problem ontology, highlighting the necessary steps to guarantee a correct and continuous human presence detection. The ontology description is then provided in a formal way. Section III presents the proposed system along with its components as a consequence of the previous analysis. Section IV reports on the experiments carried out by employing the implemented system. Section V draws conclusions.

II. Problem Ontology

The problem ontology can be defined as follows. Let's suppose an organization with the need of monitoring the presence of its staff at their personal computer during an on-line course. To access the course remotely in a private session, username and password are required. The login phase is the starting event to the proposed system detection. Generally, person identification is specifically dealt with other well-studied technologies such as for example fingerprint or iris verification, which are assumed to be known to the reader and go beyond the scope of this work. Since, in principle, each registered user is uniquely referred to a record in a database, the choice of a username and a password must be intended only as a practical solution to the problem of detecting someone's identity. This does not solve the problem of the identity stealer who purposely subtracts another's identity, which would require much more semantics. In this context a trustworthy logger is supposed.

After that, the problem ontology requires:

1. the logged person is detected as a target *human* object with a unique id
2. the detected person is *continuously* tracked by the system, eventually monitoring his/her behavior.

The first step can be accomplished specifically with image-processing techniques; the second step depends on the level of details we add to human presence description. The more chemical parameters affected by the human presence are considered, the better the tracking that can be implemented. Hereinafter an in-depth analysis about the two steps is provided.

Since the issue of identification is solved by the previous assumption concerning the login phase, the remaining problem is reduced to human face detection. The semantics of the human face is a widely debated issue [10, 11]. There are some visual features that are similar across races, genders, age, etc… constituting a sort of *skeleton* (i.e. a structural pattern) that can be revealed through image analysis techniques like edge representation [8]. Once the image features of the person to detect have been recognized by the system as being similar to the standard image skeleton of the human face, then such degree of similarity can be used to evaluate person's behavior according to simple observations. A person face staying in front of a pc monitor may stand firm, move or leave the scene, turn the head etc. and all these *behaviors* can be characterized by simple image processing. The person standing in front of pc however may be at a different distance from the image detector, then the distance parameter should be also taken into account.

Ultimately, since *human breathing* produces dioxide carbon emissions, this parameter is *ontologically* related to the human presence. Consequently, it can be used both to validate the image-based human detection mechanism and infer on simple behaviors like yawning, speaking, deeply breathing in and so on. The human presence influences the surrounding environment: i.e. it produces local changes in environmental parameters. Namely, these parameters are chemical (such as carbon dioxide emission, humidity, temperature) and physical (like object moving in the scene). In general, chemical parameters have slower response characteristics than the physical ones. This behavior must be taken into consideration when a data fusion approach among different sensors is pursued. In this context, a sensor response, whatsoever its nature is, can be considered as a *fuzzy statement* [12] describing the occurrence of a human-environment interaction event. Figure 1 displays an example. Other local parameters affected by the human presence such as humidity and temperature may be considered. The here presented problem ontology may be then enriched with several more aspects than human breathing only: however this has two pitfalls:

1. it increases the number of variables (hence sensors) needed;
2. it does not add any substantial methodological innovation to what is described here. For these reasons, the problem ontology formalization is kept as concise as possible.

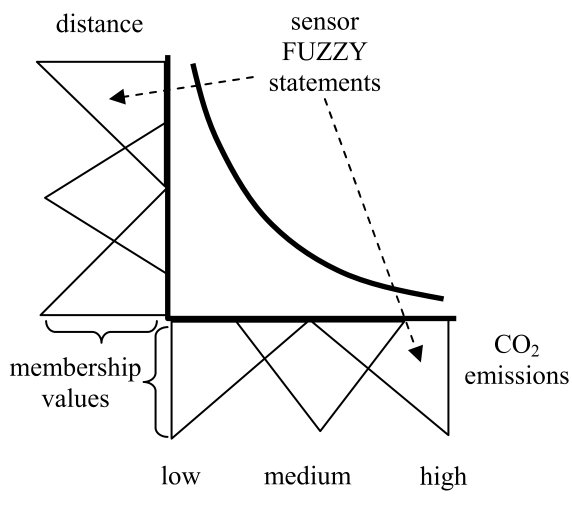

Figure 1: fuzzy description of the curve relating CO2 emission of a person to the distance from the sensor

For these reasons the following formalization describes a minimal set of assertions, but it can be easily enhanced to express the problem ontology in a more detailed way.

A. Problem ontology formalization

The above considerations can be transformed in formal logic as predicates over the set of concepts defined by the problem ontology and they are provided further on. The first three clauses account for grounding definitions, the second four clauses define different events detecting person's behaviors.

- *Detection*. Everything that performs a login and is *similar* to a standard configuration is a detected person:

  $\forall x$ login(x) $\wedge$ is_similar_to(x,'std_cfg') $\rightarrow$ detected(x)

- *Identification.* The identity of the detected person is defined by the performed login:

  $\forall x$ detected(x) $\rightarrow$ (login(x) $\rightarrow$ id(x))

- *Standard configuration.* Every standard configuration corresponds to an image skeleton taken at a certain distance:

  'std_cfd' $\rightarrow$ $\exists d$ distance(d) $\wedge$ skeleton(d)

- The detected person is standing firm in his/her position (event 0):

  $\forall$x detected(x) $\rightarrow$ (standing(x) $\rightarrow$ event(0))

- The detected persons is not standing firm in his/her position and is moving away from the place where he stays (event 1):

  $\forall x$ detected(x) $\rightarrow$ ($\neg$standing(x) $\wedge$ moving_away(x) $\rightarrow$ event(1))

- The detected person is not standing firm in his/her position and is not moving away from his place and is producing high CO2 emissions (event 2):

  $\forall x$ detected(x) $\rightarrow$ ($\neg$standing(x) $\wedge$ $\neg$moving_away(x) $\wedge$ producing_highCO2(x) $\rightarrow$ event(2))

- The detected person is not standing firm in his/her position and is not moving away from his/her place and is not producing high CO2 emissions (event 3):

  $\forall x$ detected(x) $\rightarrow$ ($\neg$standing(x) $\wedge$ $\neg$moving_away(x) $\wedge$ $\neg$ producing_highCO2 (x) $\rightarrow$ event(3))

From a modeling perspective, it is useful to split system design into a static and a dynamic view in order to enlighten both functional and algorithmic aspects. To comply with these commitments a flowchart for system working along with a component-based description are provided..

Clauses defined in the problem ontology formalization produce the flowchart depicted in Figure 2. It represents the logic steps that allow for detecting the target object (upper block) and classifying over possible events repeatedly (lower blocks) which are considered outputs for the system. It is worthy stressing that the conditional blocks do not represent boolean situations (visualized for simplicity in Figure 2 through yes-no branches); instead, they account for multi-valued conditions. For example, the block deciding if the detected person is moving or not which describes *how much* the detected person is moving. This change makes the system capable of detecting more *expressive* situations, allowing for handling vagueness like in fuzzy approaches. Theoretically each multi-valued condition can be cast to a yes-no value through appropriate threshold.

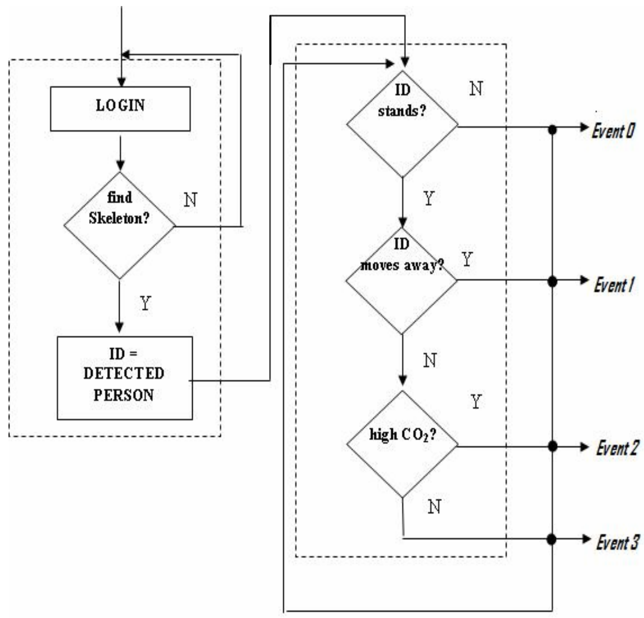

Figure 2: flowchart representing the logical steps of a simple version of the implemented human presence detection system

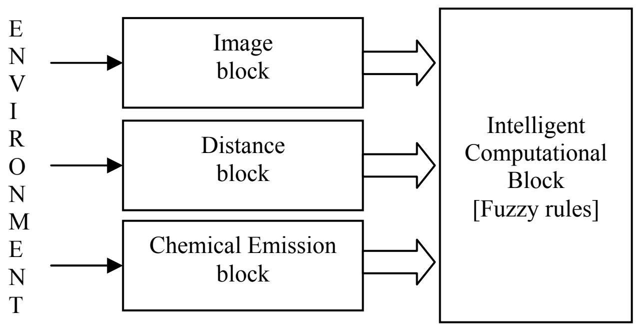

Figure 3 component-based architecture of the proposed system for human presence detection

III. SYSTEM MODELLING

In accordance with the provided problem description, a system for detecting human presence in front of pc can be designed as requiring the following components:

1. image sensor block: it consists of a digital camera for catching the scene at a given frame rate (1Hz in our experiments)
2. distance sensor block: actually an electronic distance reader;

3. chemical emission sensor block: it is composed of a set of sensors for detecting human breathe
4. Intelligent computational block: which performs signal processing at the ontology level

The Image and distance blocks account for high-response rate sensors (they detect environmental changes almost instantly) while the chemical emission block generally has lower response-rate (several seconds). The computational block, basing on fuzzy rules, mediates among the signals coming from the three sensor blocks; actually it processes signals at the ontology level and performs human detection evaluation. An outline of the abstract system architecture is provided in Figure 3.

A. Image Block

By means of a low-cost webcam, this block outputs a numerical measure concerning the likelihood between the detected object and a standard human face opportunely preprocessed. The used image analysis technique employs a low-level visual feature, called string signature [16], extracted from each video frame to represent concisely the image contained in the frame itself. In that work, the authors showed that such approach provides a significant representation of the image content that allows for extracting a string-based image description, consisting of the local contour orientation, suited for motion detection and image retrieval from a database. The original image is synthesized according to the following symbols: "/", "|", "-", "\", "o" which represent the prevalent orientation of the blocks in which the image is divided (the "o" symbol account for no prevalent orientation).In our experiments each image frame was divided into 81 blocks (9 x 9). Figure 4 shows an example.

To detect the motion as well as to search for images in the video database the distance among signatures is related to the similarity among the corresponding images. Therefore, the signatures can be clustered using well-known fuzzy clustering methods to compute the centroids of the classes, which represent the background knowledge acquired by the system in the video analysis. For the sake of simplicity one only centroid has been considered corresponding to the standard configuration (with the target face exactly in front of the webcam at a distance of about 60 cm from it). As distance measure a simple metric such as the Hamming Distance has been chosen. Since strings signatures have all the same length (81 symbols) their normalized distance against the reference string was reckoned as the number of flipping symbols divided by 81.

As working hypothesis, the following assumptions have been made:

- all objects in an image are characterized by edges (limiting a shape or constituting a boundary in the cases which are most useful for the image understanding);

Figure 4: exemplar string signature superimposed to the original image

- the target object (a human face) characterizing the semantic content of an image is in foreground.
- Every edge in an image can be approximately described by using a sequence of prevalent directions chosen in the set composed only of the four principal 2D directions (namely, horizontal, vertical, and the two diagonal ones) and by the lack of direction.

After the experiments, these constraints appeared to be not too restrictive for the problem context.

B. Distance block

This block produces a measure sensing the distance of the target object from the detector. For this aim, a low-cost sonar range detector was employed. In particular, a MaxSonar® – EZ1 sensor was chosen for experiments. In this sensor, the beam pattern is a function of the amplifier gain ranging from 3.5 V to 5 V. Beam patterns can be then calibrated in dependence on the operating conditions. In our case the mean distance of the target object from the sensor was about 60 cm. It is interesting to note that a cluttered background corresponds to several false range detections. This noise, on the other hand, carries some information since it can be generally imputed to a person moving in front of the detector as experiments show.

C. Chemical Emission block

For this block, eight low-cost chemicals sensors revealing different environmental parameters (Table 1) were considered at first stage. Sensors 1, 3, 5, 7 showed similar responses. Sensors 1, 3, 5 are affected by CO, which is present in the breathe of a smoker (used as a tester). Sensor 7 (humidity) is of course always related to human breathe. Experiments showed that sensors 4, 5 and 7 were the best candidates to achieve good performances. These three sensors were opportunely positioned into a metal jacket placed in front of the monitored area. In a very reduced and cheap configuration, these three sensors may be even substituted with a single Figaro TGS-800 sensor which detects a mixture of gaseous air contaminants. Low-cost sensors are generally sensitive to several molecules, thus making it unfeasible to discriminate the exact reagent unless more data from multiple sources are crossed. Consequently, one single sensor like TGS-800 alone is not capable of distinguishing for example between carbon monoxide and methane. Otherwise, it suggests, for the same output value, a certain concentration of CO and another concentration of CH_4. Determining the exact reagent goes beyond the scope of this work: for the aim of the paper it is sufficient to discriminate between a *good* air quality condition and a *bad* one. This is the reason why the following considerations are based on a minimal set of sensors.

TABLE 1
CHEMICAL SENSORS EVALUATED FOR EXPERIMENT PURPOSE

id	**sensor**	**Measured chemicals**
1	TGS-2442	CO
2	TGS-2201 A	NO_2, SO_2, H_2S
3	TGS-2201 B	CO , burnt hydrocarbons
4	TGS-4161	CO_2
5	TGS-2600	methane, CO, isobutane, ethanol, hydrogen
6	TGS-2100	CO, ethanol, hydrogen
7	TGS-2180	Humidity
8	TGS-800	Gaseous air contaminants

D. *Intelligent Computational block*

To fuse data at the ontology level a Fuzzy Inference System (FIS) may be considered for preliminary tests. A huge literature exists on fuzzy logic ([17], [18], [19]) and a wide variety of both academic and industrial applications is available today. Fuzzy logic allows for designing systems using multi-valued logic, hence its use makes sense whenever boolean values are not sufficient for problem description. Problem description suggests that, for the scope of this paper, the use of Fuzzy Logic is convenient. Consequently, with reference to the proposed system architecture, each sensor block output can be mapped into an FIS input. From the previous analysis, two outputs should be considered: one for human presence detection, the other for target moving. Two possible FIS rules are the following:

R1. IF string_signature_istance is *low* AND air_quality is *low* AND distance is *low* THEN human_presence is *high* and target_moving is *low*

R2. IF string_signature_istance is *high* AND air_quality is *low* AND distance is *high* THEN human_presence is *high* and target_moving is *high*

The first rule accounts for a very desirable condition, while the second one attempts to catch out-of-normal situations imputable to a moving target. In these rules air quality information is slightly privileged against other sources since it is strongly ontologically related to human breathe (although other parameters may affect air quality as well).

Figure 5: Multi-view image of the experimental device for sensing human presence comprising the distance block and the chemical emission block of Figure 1. The left mage illustrates the whole handcrafted device made of a LabJack UE9 (data acquisition unit with Ethernet interface) and a cardboard support for sonar range detection and air quality sensor. In the right image, the sonar range detector is visible in the forefront. A fan (foreground) inside the cardboard jacket is used to convey air towards the air quality sensor (behind sonar range detector).

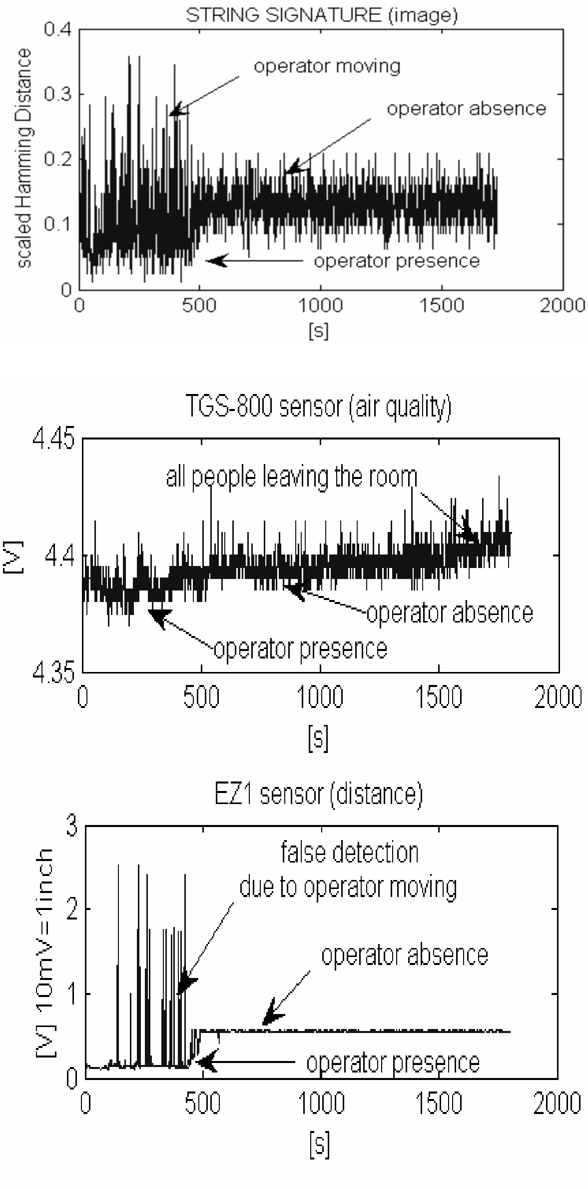

Figure 6: sensor responses in a minimal configuration
(one sensor per block)

IV. EXPERIMENTS

In this section a first glance over the robustness of the proposed approach is provided through experiments on real scenarios.. In accordance with the component-based architecture depicted in Figure 3, a multi-sensor equipped device has been experimentally setup to conduct measurements on real scenarios. A preview of the implemented system comprising distance block and chemical emission block with one sensor each is presented in Figure 5. For experiment purpose, a person has been asked to be present in front of the system for a given amount of time and then to swiftly leave. This action has been repeated over time with progressively lesser presence/absence interval time in order to assess the minimal response frequency of chemical sensors. Results showed that with interval times less than about 5 minutes the system rapidly degrades the performance due to slow response time of the used sensors. In this case, however, the signal coming from the sonar range detector characterized by extremely high response time (approximately 50ms) may be given a higher weight in the presence/absence decision process. An agent-based mechanism to handle trustworthy among sensors may be hypothesized but goes beyond the scope of this paper. With reference to Figure 6, it is interesting to note the correspondence between FIS rules and real world conditions. After preliminary experiments we found that the provided description was able to detect human presence with high success rate (generally more than 85%). Our first experiments showed that the proposed ontology-based approach is effective for handling complex contexts like human presence detection; a wider experiment set and a fine tuning of the FIS architecture will be considered in future research on the subject. Nevertheless, the authors are confident that data-fusion problems related to such contexts will take benefit from this approach.

V. CONCLUSION

In this work, an ontology-based approach to automated human presence detection has been presented. The ontology description led the authors to hypothesize a two-level architecture composed of sensor blocks for catching images, distances and chemical emissions of the target and finally an intelligent computational block that supports data fusion at the ontology level. First tests confirm that data heterogeneity and complexity can be managed better if handled at the ontological level by means of a FIS. Laboratory experiments obtained a success in detecting human presence with a good degree of accuracy until inertia related to (chemical) sensor slow response rate becomes prominent. It is therefore evident that the adoption of high response rate sensors like a webcam and a sonar range detector represents a solution to this problem which deserves however a specific analysis and will be handled in future works. Our prospective research will be also aimed at selecting the best combination of sensors to achieve higher performances.

ACKNOWLEDGMENT

This work was supported in part by the EU, within the 7FP project, under grant agreement 216483 “PrimeLife”

REFERENCES

[1] Masahiro Nishi Shigeru Takahashi Teruaki Yoshida , Indoor Human Detection Systems Using VHF-FM and UHF-TV Broadcasting Waves. IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications, pp. 1-5 , 2006.

[2] Nanzer, J.A. Rogers, R.L., Human Presence Detection Using Millimeter-Wave Radiometry, IEEE Transactions on Microwave Theory and Techniques, Volume: 55, Issue: 12, Part 2, pp.: 2727-2733, Dec. 2007.

[3] Jianchao Zeng; Sayedelahl, A.; Chouikha, M.F.; Gilmore, E.T.; Frazier, P.D., Human detection in non-urban environment using infrared images 6th International Conference on Information, Communications & Signal Processing, pp.:1 - 4, 10-13, Dec. 2007.

[4] Nanzer, J.A. Rogers, R.L., A Ka-band Correlation Radiometer for Human Presence Detection from a Moving Platform, IEEE/MTT-S International Microwave Symposium, pp.: 385-388, 3-8 June 2007.

[5] Jianpeng Zhou Jack Hoang, Real Time Robust Human Detection and Tracking System IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, pp.: 149-149, 25-25 June 2005.

[6] Nishi, M.; Kawaguchi, T.; Takahashi, S.; Yoshida, T., Human detection system using UHF band terrestrial TV receiving waves IEEE Antennas and Propagation Society International Symposium, pp.:3097 - 3100, 9-14 July 2006.

[7] V. Di Lecce, M. Calabrese, Experimental System to Support Real-Time Driving Pattern Recognition, Lecture Notes in Computer Science, Vol. 5227/2008, pp. 1192-1199, Springer Berlin / Heidelberg (2008).

[8] X. Chen, T. Delvecchio, V. Di Lecce, Deriving Semantic from Images Based on the Edge Information, Proceedings of 15th Conference on Information Modelling and Knowledge Bases.Tallinn, Estonia, May 15-19 pp. 260-267, 2005.

[9] Hall, D.L. Llinas, J., An introduction to multisensor data fusion, Proceedings of the IEEE, Vol. 85, N. 1, Vol.: 85 , Issue: 1, pp.: 6 – 23, Jan. 1997.

[10] Adler, A. Schuckers, M.E., Comparing Human and Automatic Face Recognition Performance, IEEE Transactions on Systems, Man, and Cybernetics, Part B, Volume: 37 , Issue: 5, pp.: 1248 – 1255, Oct. 2007

[11] De Vel, O. Aeberhard, S., Line-based face recognition under varying pose, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.: 21 , Issue: 10, Oct. 1999.

[12] Mendel, J.M., Computing with Words: Zadeh, Turing, Popper and Occam, IEEE Computational Intelligence Magazine, Vol. 2, Issue 4, pp.:10 – 17, Nov. 2007.

[13] Felix Opitz, Werner Henrich, Thomas Kausch, Data Fusion Development Concepts within complex Surveillance Systems, Fusion, 2004.

[14] Dastner, K.; Kausch, T.; Opitz, F., An object oriented development suite for data fusion: Design, generation, simulation and testing, 10th International Conference on Information Fusion, pp.:1 – 8., July 2007.

[15] Hills, J.W. Jensen, J.F., Telepresence technology in medicine: principles and application, Proceedings of the IEEE, Vol. 86, Issue: 3, pp.: 569-58, Mar 1998.

[16] A. Amato, V. Di Lecce, Edge detection techniques in image retrieval: The semantic meaning of edge, EC-VIP-MC 2003, 4th EURASIP Conference, Zagreb - Croatia, pp. 143-148, July 2-4 2003.

[17] Zadeh L., Knowledge representation in fuzzy logic, IEEE Transactions on Knowledge and Data Engineering, Vol. 1, pp. 89-100, 1989.

[18] Bezdec, J.C., Pattern Recognition with Fuzzy Objective Function, Plenum Press, New York (1981).

[19] Mamdani E.H. and Assilian S., An experiment in linguistic synthesis with a fuzzy logic controller, International Journal of Human-Computer Studies, Volume 51, Number 2, pp. 135-147, 1999.