Autonomous perception techniques for urban and industrial fire scenarios.

J. Capitán, D. Mantecón, P. Soriano and A. Ollero
Robotics, Computer Vision and Intelligent Control Group
Escuela Superior de Ingenieros, Camino de los Descubrimientos s/n, 41092 Sevilla, Spain
[jescap, mantecon, psoriano, aollero]@cartuja.us.es

Abstract—This paper presents autonomous perception techniques for disaster management and particularly for urban and industrial fire scenarios. The perception from static and mobile (on-board UAV) cameras is considered for smoke detection and cooperative tracking The algorithms used to achieve these functionalities will be described. The implementation of these techniques in the AWARE platform and the application in the general AWARE experiments to detect smoke and track firemen of the Seville fire brigades will also be presented.

Keywords**:** *Computer vision, Cooperative Perception, fire scenarios, smoke detection, person tracking.*

I. Introduction

This paper deals with the development of autonomous perception functionalities to be used in disaster management scenarios. Particularly, the paper considers two different valuable functionalities in urban and industrial fires: smoke detection and tracking of persons.

Disaster management scenarios usually deal with dynamic environments and varying conditions for perception. The robustness and reliability of autonomous perception in these scenarios are main issues. In many cases a single autonomous entity (i.e. robot or camera) is not able to acquire all the information required for the application because of the characteristic of the particular task or the harmful conditions (i.e. loss of visibility). Then, in these scenarios, the cooperation of several of these entities is relevant.

The work described in this paper has been carried out in the European Commission project AWARE on the Autonomous self-deploying and operation of Wireless sensor-actuator networks cooperating with Unmanned Aerial vehicles (UAVs).

Although the project consists of the development of the whole platform and involves issues related to the cooperation and communication among the different UAVs and Ground Sensor Networks (GSNs), this paper focuses on the Perception System (PS). This system is based on building and updating a consistent representation of the environment suitable to achieve detection and tracking using the sensors provided by the AWARE platform: visual and infrared cameras, sensors of scalar magnitudes such as temperature, humidity, CO or radio signal strength. The AWARE PS system has been designed to detect events such as fires and to perform cooperative tracking of firemen and vehicles in the fire scenario by using the information from Ground Camera Nodes (GCN), cameras mounted on autonomous helicopters, and nodes of the wireless sensor network (the firemen also carry nodes of this network). Figure 1 illustrates the tracking.

A distributed architecture has been adopted for the PS in AWARE. This architecture decreases the requirements on data transmission and improves the scalability of the whole system. In addition, it will be possible to divide the processing load among different perception nodes. These nodes will process locally the environment information (images, sensor readings, ...) in order to decrease the amount of data transferred through the network. All the nodes will share the perception information extracted from their local sensors and then, improve its knowledge of the environment by integrating the measurements from other nodes. AWARE will be demonstrated on fire detection and monitoring in urban scenarios, and on tracking the position of people and vehicles. This paper presents results on autonomous smoke detection, which is a basic functionality in urban fires. Furthermore, the paper presents results on the cooperative tracking of firemen. These results have been obtained in the first general experiments of AWARE carried out in March 2007.

The paper is organized as follows; Section II describes perception techniques developed to achieve the smoke detection functionality. Section III and IV detail the firemen detection and cooperative tracking algorithms respectively. Section V is devoted to present some experiments. Finally, section VI gives some conclusions and future work.

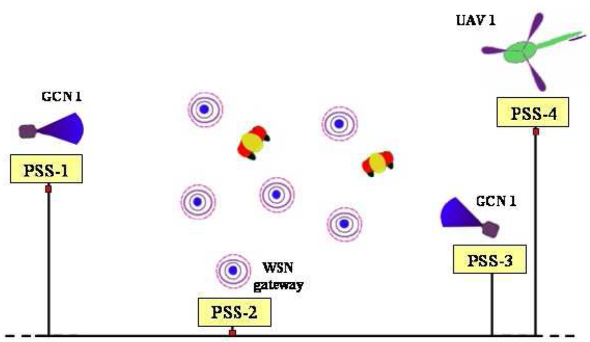

Fig. 1. Tracking of persons using ground cameras, UAV and nodes of a wireless sensor network.

II. VISUAL BASED SMOKE DETECTION

Autonomous fire detection in outdoor environments has been a research and development subject in the last decades. Several infrared detection techniques, including false alarm rejection have been developed. The autonomous detection of the smoke is attractive because of the today low cost and low maintenance of visual cameras. However, several authors (see for example [1], [2]) have shown the difficulty of smoke segmentation in images using classical methods due to the lack of discriminating features capable of differentiating between smoke and background in all cases. Smoke features such as color, size or speed of spread depend on many parameters such as the material burnt. Moreover, the wind direction and speed can change, even for the same fire, if the parameters change. Besides, smoke has no edges, corners or defined shape due to its non-rigid gaseous nature. In general, it is not possible to find a single feature capable of segmenting smoke successfully. The approach proposed in this paper is not to use a single discriminating feature but to exploit the synergies between different features. Then, several features detectors have been implemented and their outputs are later combined within a probabilistic framework by a fusion data layer to reduce false alarms.

The approach can be divided in two steps. In a first step, a set of features are extracted from the images. A bank of algorithms extracts features such as independent color, shape, texture and motion. The outputs of these algorithms will be viewed as Bernoulli random variables, in such a way that each detector measures whether there is or not an alarm (smoke) for each pixel. In a second step, these outputs are combined in an occupancy grid by a global estimator (see figure 2). This type of filtering has been successfully employed in earlier works such as [3] and [4]. The overall estimation is expected to be more reliable to variations of single features. For the integration, the smoke features are considered independent. It is expected that, where one of the features fails (i.e., give a false positive, or a false negative) another will succeed, making it possible to balance the results. Up to now, three main smoke features are considered: color, motion and texture. Future implementations will consider blob analysis and background learning.

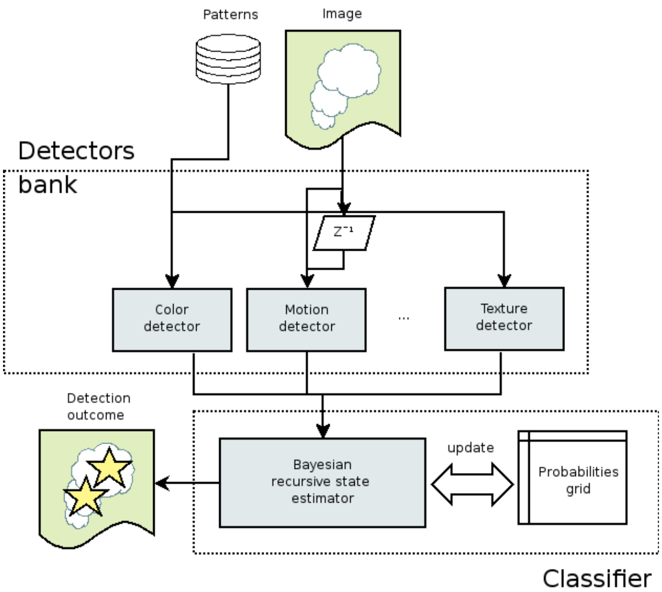

Fig. 2. Block diagram of multi-feature image based smoke detection

A. Color Detection

Smoke color cannot be represented accurately by a single unimodal probability density function (p.d.f.). Its color depends on the illumination conditions and also on the material that is being burnt. Thus, to accommodate different ranges of color, p.d.f. with multiple modes (i.e. multimodal p.d.f.) are required. A statistical classifier that models the classes as probability density functions has been developed. Particularly, a particle filter, able to cope with multimodal p.d.f. modeling has been implemented. Furthermore, likelihood has been chosen as the metric, under 3D work space (RGB or HSI).

For an efficient measurement of the distance, the three color components are considered as independent random variables. Thus, the 3D joint probability density function can be decomposed in three marginal unidimensional distributions over each color axis. Three Gaussian mixtures (one per color) are built from the particles in the filter. The above described technique can be particularized for different 3D color representation such as RGB and HSI.

B. Motion Detection

Smoke spread and motion can be of interest for its identification and segmentation from the background. In the AWARE project, three basic approaches for motion analysis have been addressed: feature identification and tracking, optical flow and image subtraction techniques.

The detection and tracking of features such as corners or edges in the images is not valid for smoke segmentation. First, it is not easy to identify clear features in smoke images. Second, the features that were identified can not be tracked easily due to the gaseous nature of smoke. Optical flow techniques have similar problems: smoke does not present features consistent in time.

Image subtraction techniques can be applied only in case the camera and scene are static. In the case of moving cameras, it is necessary to stabilize the images, which can be also done by applying image processing techniques. Before subtracting, the images should be filtered to avoid the influence of noise. A Gaussian blur low-pass filter has been tested with successful results. The result of the image subtraction is smoothed using a moving average time filter.

C. Texture Detection

The basic idea of Texture Unit (TU) methods [5] is to apply a transformation to every pixel in the image in such a way that, in the new domain, each pixel value depends on the contrast variations in its neighborhood. In the original work, the transformations have to threshold the eight neighborhood of a pixel against the central value, and assign a ternary value (0, 1 or 2) depending on a parameter d, see [5]. The number generated by composing the eight ternary digits shows a histogram that remains stable over the regions on the image with similar texture.

A different technique, Local Binary Pattern (LBP [5]), has demonstrated to be more appropriate for the problem. In LBP each pixel of an eight-neighborhood is binarized after comparing with the value of the central pixel. The result of the binarization is considered as the bits of an 8-bit integer in the range [0, 255]. This integer value (one per pixel) takes into account the local variations of the levels in the original image.

LBP is independent from the illumination conditions of the image. This method is very sensitive to relatively small variations in the gray level intensities. Smoke has smother texture than many other natural objects in the image such as trees and ground.

III. Firemen Detection Using Visual Images

Autonomous detection and tracking of people by means of visual image sequences is a well known subject, but there are still many open issues depending of the difficulty of the scenario considered (occlusions, number of moving objects, …). Thus, different features and constraints are considered to discriminate people from other moving objects in the sequence of images. Thus, some works ([6], [7], [8]) consider the verticality of people's edges, the typical people's sizes and shape, and even the model of walking. However, in most cases, adopted solutions only are expected to be useful under certain conditions or in specific scenarios. This section describes the techniques used to achieve the correct segmentation and localization of the firemen throughout an image sequence. The consideration of firemen simplifies the problem because it is possible to track using mainly color features However, learning techniques could be also applied to track designated people.

The technique described in this section will be used independently for each visual source and will provide the necessary measurements for the cooperative tracker.

A general overview of the firemen identification method is shown in the figure 3. It applies color segmentation and blob segmentation to identify coherent regions in the image. Both features are inputs of a detection decision system. Finally, when a fireman is detected, its position on the image plane is estimated. These measurements will be used as inputs for the cooperative tracker.

A. Color Segmentation

The color features of the particular firemen uniform are used to train off-line the classifier used for segmentation. .

Segmented pixels are grouped in clusters. Each cluster represents a set of nearby pixels with the desired color and will be interpreted as a separate object, in this case a fireman. This technique allows to segment several objects in the same image, and thus to perform tracking of multiple objects. Some

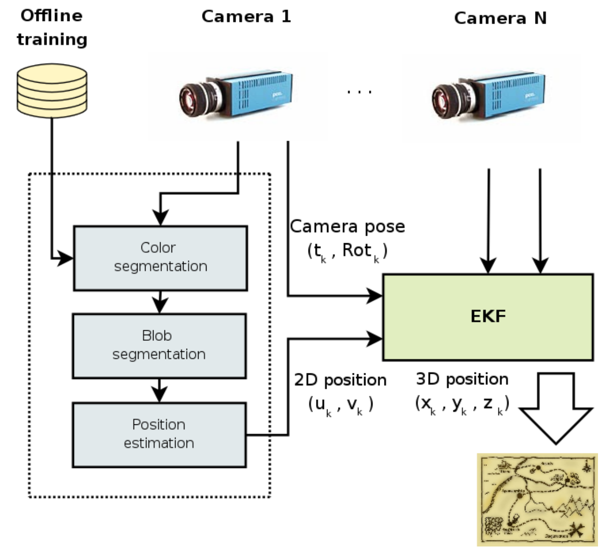

Fig. 3. Block diagram of the cooperative tracking system

clusters may be originated by spurious objects and other by desired objects. The clusters are selected by eliminating those that do not fulfill a set of constraints related to their shape, minimum number of pixels, etc. Then characteristics of the resulting clusters will be calculated: each of them is a potential fireman in the image.

B. Blob Segmentation

The objective is to check if there is a region of homogeneous color around each color cluster. Blob extraction does not attempt to segment out exact shapes of objects. Instead it tries to extract robust and repeatable features, easier to track. Search areas are defined around the cluster defined by the color segmentation method. Blob extraction method will be applied in this search area. There are a wide variety of blob extraction techniques such as [9] or [10]. The technique used in this paper is based on [10].

C. Fireman Detection

Finally, the objects, in this case firemen, are detected from the information of the blob extraction and color clusters. A classifier will differentiate between firemen and other objects using information and scores obtained from blobs and color clusters. If there is a blob near a cluster center, whose properties and moments correspond to a homogeneous region with vertical shape and similar color to the uniform, it will be considered a fireman. In this case, the centroid of the blob will be provided to the tracker as a measurement in the image plane.

IV. Cooperative Tracking

The probabilistic framework in AWARE allows the integration of different sources of measurements from different cameras to decrease the uncertainties in the estimated 3D localization, which is the final output of the tracker. Different techniques can be implemented in this framework. Although the AWARE perception system architecture is distributed, the first implementations of the cooperative tracking in the AWARE project is based on a single Extended Kalman Filter (EKF) ([11]) which receives measurements from all the available visual sources. So, once the segmentation and detection algorithms have been done in each camera independently, measurements on the image plane are sent to the EKF which is in charge to integrate them in order to provide the 3D localization of the tracked fireman.

A. The state vector

The state to be estimated consists of the 3D position and velocity of a person, both in the global coordinate system.

$$\mathbf{x}_k = \begin{bmatrix} \mathbf{r}_k \\ \mathbf{vel}_k \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \\ vel_x \\ vel_y \\ vel_z \end{bmatrix} \quad (1)$$

B. Prediction Model

The sample time is assumed to be small enough to approximate the person motion as linear. A random and gaussian variation (mean zero) for the velocity vector is considered at each step. Thus, the linear equations for the prediction step and the corresponding noise vector are:

$$\mathbf{\Delta vel} = \begin{bmatrix} \Delta vel_x \\ \Delta vel_y \\ \Delta vel_z \end{bmatrix} = \begin{bmatrix} N(0,\sigma_{vel}^2) \\ N(0,\sigma_{vel}^2) \\ N(0,\sigma_{vel}^2) \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{r}_k \\ \mathbf{vel}_k \end{bmatrix} = \begin{bmatrix} \mathbf{r}_{k-1} + (\mathbf{vel}_{k-1} + \mathbf{\Delta vel})\Delta t \\ \mathbf{vel}_{k-1} + \mathbf{\Delta vel} \end{bmatrix} \quad (2)$$

C. Measurement Model

The measurement vector is composed by the location and velocity of the person referenced to the image plane, expressed in pixel and pixel/*s* respectively. Measurement uncertainties are considered gaussian again and they will depend on the segmentation algorithm accuracy.

$$\mathbf{z}_k = \begin{bmatrix} \mathbf{z}_{k,1} \\ \mathbf{z}_{k,2} \end{bmatrix} = \begin{bmatrix} u_k \\ v_k \\ \frac{u_k - u_{k-1}}{\Delta t} \\ \frac{v_k - v_{k-1}}{\Delta t} \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} \quad (3)$$

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} N(0,\sigma_u^2) \\ N(0,\sigma_v^2) \\ N(0,(\frac{\sqrt{2}}{\Delta t}\sigma_u)^2) \\ N(0,(\frac{\sqrt{2}}{\Delta t}\sigma_v)^2) \end{bmatrix} \quad (4)$$

In order to model the transformation from the state vector to the measurement vector, the global 3D location is expressed in the camera system coordinates and then projected to the image plane by means of a calibration matrix. This projection involves a non-linearity due to a scale factor (*s*), and then the resulting vector is divided by the last component to be homogenized.

Let be $\mathbf{A_{cal}}$ the calibration matrix, ${}^{w}_{c}\mathbf{Rot}_k$ the rotation matrix from the world coordinate system to the camera one, and $\mathbf{t}_c$ the translation vector of the camera coordinate system respect to world coordinate system. Then, the equations for the measurement step follows:

$$\mathbf{z}_{k,1} = \begin{bmatrix} u_k \\ v_k \end{bmatrix} = \frac{1}{s} \cdot \mathbf{A}_{cal} \cdot \begin{bmatrix} {}^{w}_{c}\mathbf{Rot}_k^T & -{}^{w}_{c}\mathbf{Rot}_k^T \cdot \mathbf{t}_{c,k} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{r}_k \\ 1 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \quad (5)$$

$$\mathbf{z}_{k,1,prev} = \frac{1}{s'} \cdot \mathbf{A}_{cal} \cdot \begin{bmatrix} {}^{w}_{c}\mathbf{Rot}_{k-1}^T & -{}^{w}_{c}\mathbf{Rot}_{k-1}^T \cdot \mathbf{t}_{c,k-1} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{r}_k - \Delta t \cdot \mathbf{vel}_k \\ 1 \end{bmatrix} + \begin{bmatrix} v_{prev,1} \\ v_{prev,2} \end{bmatrix} \quad (6)$$

$$\mathbf{z}_k = \begin{bmatrix} \mathbf{z}_{k,1} \\ \mathbf{z}_{k,2} \end{bmatrix} = \begin{bmatrix} \mathbf{z}_{k,1} \\ \frac{\mathbf{z}_{k,1} - \mathbf{z}_{k,1,prev}}{\Delta t} \end{bmatrix} = \begin{bmatrix} u_k \\ v_k \\ vel_{u,k} \\ vel_{v,k} \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} \quad (7)$$

The final noise vector of the measurement step ($\mathbf{v}_k$) is composed by the four additive noises described above and the rotation and translation matrices uncertainties.

The General EKF equations are used to perform the filter steps with the models described above.

D. Filter initialization

Within the framework of the AWARE project, Wireless Sensor Networks (WSN) are also being used for detection and tracking. Firemen are supposed to carry a WSN node and the position of the rest of the nodes are assumed to be known as well. Applying stochastic filtering and received signal strength on the WSN nodes, a first localization of the fireman can be obtained. EKF state will be initialized from this WSN estimation.

V. EXPERIMENTS

The experiments took place in Utrera (near Seville, Spain) in March 2007 with the participation and assistance of the Fire Brigades of Seville.

A. Smoke detection

This experiment was carried out using a smoke generation machine. A three-floor building was simulated with a structure and the smoke machine placed in its second floor. The smoke was filmed from both ground and aerial cameras and the sequences processed with the proposed methods. The overall classifier was applied in order to fuse the results provided by the different detectors. Results showed to be very good and smoke was correctly segmented from the original images. Outputs of the individual detectors are shown in figure 4. Smoke was detected correctly for a very high percentage of the cases and most of the false alarms provoked by reflections in the images were discarded thanks to the combination of the three detectors. The smoke detection has been implemented for both the fixed Ground Camera Nodes and the camera on board the Flying Cam Helicopter. Videos illustrating these detectors will be presented in the conference.

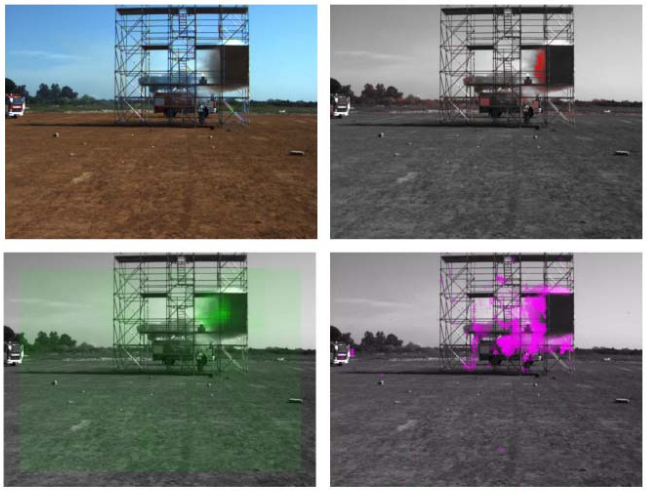

Fig. 4. From left to right and from up to down, the original image, color-based detected smoke, texture-based detected smoke and motion-based detected smoke

Fig. 5. Flying Cam helicopter during the experiments

B. Tracking

This experiment consisted of the detection and tracking of a fireman by processing visual images. The uniform of the fire brigade of Seville is composed by blue dark trousers and a blue dark jacket with fluorescent yellow strips. Fluorescent yellow is not common in a typical scene. The binary image obtained from color segmentation classifies the pixels with the trained color (yellow fluorescent) in black and the rest of pixels of the image background, in white.

A fireman walked around a field of landmarks while he was filmed simultaneously with several cameras. Two ground camera nodes (GCN) were fixed to film from different angles the positions were measured with a GPS and their orientations where also measured. In order to get an aerial image sequence, a helicopter from Flying Cam flew (see figure 5) over the area filming the same fireman.

Orientation and GPS position of the helicopter were provided during the whole experiment and were able to be used as inputs for the tracker. The position and orientation of the GCNs were equally used as inputs. All the cameras taking part of the experiment were previously calibrated.

The image segmentation techniques proposed in this paper were used to localize the fireman throughout the different sequences, which were processed separately. Color segmentation and clustering were applied in order to create search areas where blob segmentation was performed. Good results were obtained for aerial and ground sequences after training the color detector with them. Some of these results are shown in figure 6.

The cooperative tracking algorithm fused all the measurements and provided a 3D trajectory. Despite the fact that it does not exist a real ground true for the experiment, landmarks positions were measured with the GPS so that the real trajectory followed by the fireman could be compared later with the one estimated by the tracker. Therefore, all the GPS positions of the landmarks, the GCNs and the building perimeter where expressed in the UTM system (meters) and were represented together with the estimated trajectory of the fireman provided by the cooperative tracker. The results for these experiments are shown in figure 7.

For the presented experiment the estimated trajectory was very close to the real one. An approximation of the real

Fig. 6. On the left, original image. On the right, the segmented blobs

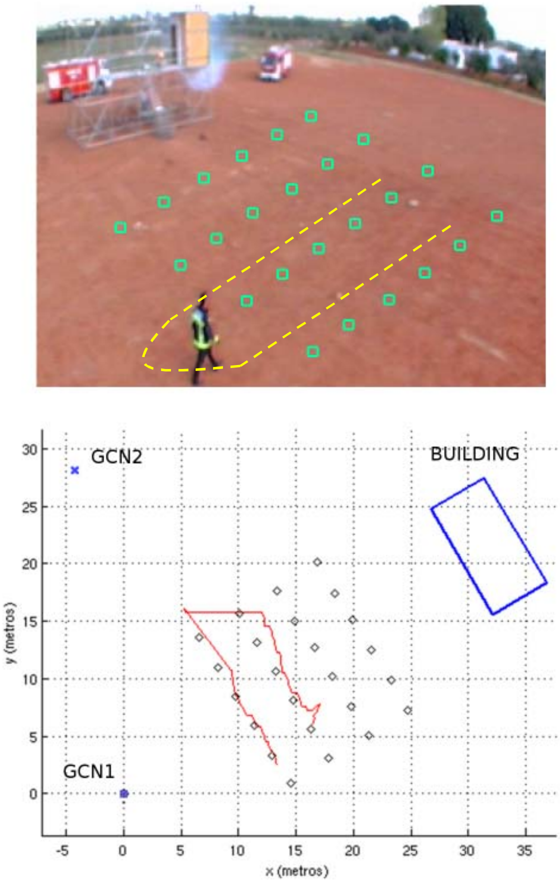

Fig. 7. Up, real fireman trajectory. Down, fireman trajectory estimated by the cooperative tracker

trajectory followed by the fireman is shown in figure 7. Furthermore, standard deviations for the 3D position were less than a meter. During the experiment, the fireman went out of the field of landmarks, getting out of the GCN1's field of view during some seconds, consistently, the estimation uncertainties increased during this time interval, and when the fireman came back to the image, uncertainties went down again. The videos to be shown in the conference will illustrate this experiment

VI. Conclusions

This paper has presented perception techniques to be applied in urban and industrial fires scenarios. In particular, smoke detection techniques and cooperative tracking have been presented. The paper includes results from the first AWARE experiments (Utrera 07).

The proposed smoke detector has demonstrated the ability to detect autonomously fires in different conditions. The cooperative tracking system also demonstrated efficiency in firemen tracking. Although a centralized cooperative system has been presented, the next step is the implementation of a fully distributed system.

Future work will also aim to fuse the methods proposed in this paper with the processed information from the Wireless Sensor Network.

Acknowledgement

The authors would like to thank the cooperation of the Seville Fire Brigades which played an important role in the performance of the experiments described in this paper. The authors also thank the contribution of the partners of the AWARE project. This project is funded by the European Commission under the contract IST 2006-33579.

References

[1] J. Chang, D. Edwards, Y. Yu. "Statistical estimation of Fluid Flow Fields", 2002, ECCV workshop on Statistical Methods in Video Processing.

[2] D. Cremers and F. Soatto. "Dynamic Texture Segmentation", 2003, International Conference on Computer Vision.

[3] L. Merino, F. Caballero, J. R. Martinez, J. Ferruz, and A. Ollero. "A cooperative perception system for multiple UAVs: Application to automatic detection of forest fires", Journal of Field Robotics, 2006, 23(3):165-184.

[4] T. Schmitt, R. Hanek, M. Beetz, S. Buck, and B. Radig, "Cooperative probabilistic state estimation for vision-based autonomous mobile robots", Transactions of Robotics and Automation, 2002, 18:670-684.

[5] T. Ojala, M. Pietikainen, and D. Harwood. "A comparative study of texture measures with classification based on feature distribution", Pattern Recognition, 1996, 22(1):51-59.

[6] T. Zhao, R. Nevatia and Fengjun Lv. "Segmentation and Tracking of Multiple Humans in Complex Situations", Proceedings of the 2001 IEEE Computer Society Conference on CVPR.

[7] T. Zhao R. Nevatia. "Tracking Multiple Humans in Crowded Environment", Proceedings of the 2004 IEEE Computer Society Conference on CVPR.

[8] Nakazawa Atsushi, Kato Hirokazu, Hiura Shinsaku, Inokuchi Seiji. "Tracking Multiple People using Distributed Vision Systems", Proceedings of ICRA, 2002.

[9] M. Donoser, B. Horst, and M. Wiltsche. "Color blob segmentation by mser analysis", In Proceedings of International Conference on Image Processing, 2006, pp 757-760.

[10] Per-Erik Forssén. Low and Medium Level Vision using Channel Representations. PhD thesis, Linköping University, 2004. Thesis No. 858.

[11] S. Thrun, W Burgard, and D. Fox. Probabilistic Robotics. The MIT Press, 2005.