Multimodal monitoring of cultural heritage sites and the FIRESENSE project *

Albert Ali Salah
Boğaziçi University
Dept. of Computer Engineering
34342 Bebek, Istanbul, Turkey
salah@boun.edu.tr

Jungong Han
Centrum Wiskunde & Informatica
Science Park 123, 1098XG
Amsterdam, the Netherlands
jungong.han@cwi.nl

Eric Pauwels
Centrum Wiskunde & Informatica
Science Park 123, 1098XG
Amsterdam, the Netherlands
eric.pauwels@cwi.nl

Paul de Zeeuw
Centrum Wiskunde & Informatica
Science Park 123, 1098XG
Amsterdam, the Netherlands
paul.de.zeeuw@cwi.nl

ABSTRACT

The FIRESENSE FP7 project aims to implement an automatic early warning system to remotely monitor areas of archaeological and cultural interest from the risk of fire and extreme weather conditions. This challenging task requires the operation of a multimodal wireless sensor network, the setting up of an infrastructure to publish and access sensor data and the fusion of multiple modalities in a real-time fashion. This paper discusses the multimodal sensor data access and fusion aspects of the project.

Categories and Subject Descriptors

I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding; I.5.4 [**Pattern Recognition**]: Applications—*computer vision, signal processing*; I.4.3 [**Image Processing and Computer Vision**]: Enhancement—*registration*; I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis—*sensor fusion*

General Terms

Algorithms

Keywords

Multimodal fusion, smoke detection, image registration

1. INTRODUCTION: THE FIRESENSE PROJECT

Cultural heritage sites are long-cultivated treasures that need special care and protection against elemental, natural and human-induced risks. In particular, fire and extreme weather related catastrophes can cause irreparable damage in a very short time. Since these areas are often surrounded by nurtured trees and vegetation, the risks due to fire are even greater.

FIRESENSE is a Specific Targeted Research Project of the European Union's 7^{th} Framework Programme (FP7-ENV-2009-1-244088-FIRESENSE)[1]. Its purpose is described by the full project name: Fire Detection and Management through a Multi-Sensor Network for the Protection of Cultural Heritage Areas from the Risk of Fire and Extreme Weather Conditions. It aims to implement an automatic early warning system to remotely monitor, through specialized sensor networks, areas of archaeological and cultural interest from the risk of fire and extreme weather conditions [6, 8]. The FIRESENSE consortium consists of 10 partners; six academic, three SMEs and one state authority, respectively. The project runs between Dec. 2009 - Dec. 2012.

Within its general scope, FIRESENSE tackles several challenges to deliver a maximally useful working system. The first and foremost of these challenges is to detect fire and smoke in a timely manner. Different sensors are employed in tandem for this purpose. While local weather information can be used to assess the risk of fire due to natural causes, human factors are not negligible. Subsequently, large areas are continuously monitored by the installed systems for evidence of fire and smoke.

The general architecture of the FIRESENSE system is given in Fig. 1. Visible range and infrared cameras are used to monitor the area under surveillance from a high position. Additionally, on-site sensors that measure temperature, humidity, CO-levels, and other parameters, as well as local weather data are gathered. Based on these modalities, the data fusion module produces an alarm level. The usability of the system crucially depends on the high true positive rate, as well as the low false positive rate of the alarms. As we

ISABEL'11, October 26-29, Barcelona, Spain.

[1]http://www.firesense.eu

will show later, multimodal fusion can help reduce the false positive rate by employing additional sensor data streams.

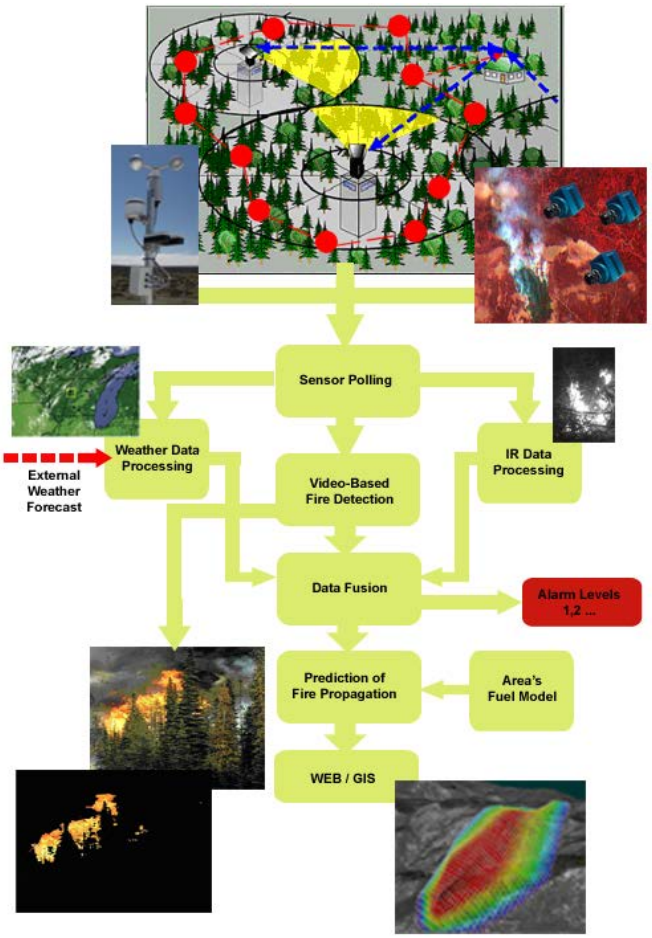

Figure 1: The overall Firesense system architecture.

The most important part of the FIRESENSE project is accurate detection of the wildfire. There are multiple methods proposed in the literature for flame and smoke detection from video input. Signal processing methods have been proposed to detect fire from video input in [21]. A smoke detection algorithm for monitoring forest fires was proposed in [7]. Calderara *et* al. recently proposed a Bayesian approach to detect smoke by analyzing image energy, but our experiments with it showed that the proposed approach is not suitable for distant events observed with a low-resolution camera [3].

In the current FIRESENSE setup, we use the smoke detection algorithm proposed by Habiboglu *et* al. [9]. This method uses statistical background models and color thresholds in conjunction to find regions in the camera's field of view that are both smoke colored and slowly moving. Spatio-temporal blocks are taken from these regions, and correlation features are extracted. The decision of smoke vs. no-smoke is given by a binary support vector machine classifier. While this system has excellent true positive rates, wind-induced camera motion may cause increases in false positive rates. This problem is tackled in Section 3.2. An IR-based fire detection system to be used in conjunction with the visible camera-based system is under development.

The second part of the system, to be used in the event of a fire, concerns the prediction of fire propagation. Since the parameters of the monitored area is well-known in advance, area fuel models, topology, weather information, as well as real-time satellite information can be employed in this part of the system. Also, a 3-D Geographic Information System (GIS) environment is used to visualise predicted fire propagation. In this paper, however, we focus on the fusion part of the system. Section 2 summarizes the software infrastructure for accessing sensor data. Section 3 details two multimodal sensor fusion components that are used in the system. Finally, Section 4 concludes the paper.

2. ACCESS TO SENSOR DATA

The design of the FIRESENSE system places a large number of sensors around the monitored location. To simplify the integration of new sensors to this system, and to provide a uniform and accessible interface to the data streamed by these sensors, we use a data publishing approach. Some web-based services like Pachube[2] offer simple and straightforward ways of publishing a real-time sensor data streams on the web. However, Pachube does not enforce strict meta-data annotation procedures, which of course can be considered a reasonable requirement for the goal of simple data publishing. However, the FIRESENSE system involves multiple types of sensors, deployed to heritage sites in different countries, and interoperability is important for both modules and sensor streams. We therefore opt for a more structured alternative.

The OpenGIS Sensor Observation Service Interface Standard (SOS) is an approved Open Geospatial Consortium (OGC) standard[3], which defines a web service interface for publishing and reading sensor data. It supports sensor discovery, the possibility of querying real-time or archival data, and works with different types of sensors. Meta-data related to sensors, like positions with respect to a given frame of reference, measurement units, calibration information, and ownership, can be stored and retrieved in a standard way. While OGC has more standards to publish data streams, including Web Feature Service (WFS) and Web Coverage Service (WCS), SOS is determined to be a better model, and is more suitable [1]. 52°North offers an Open Source implementation of this service, as well as client applications to access and visualize data streams published via SOS[4].

In our implementation, the actual sensor data are stored in specifically structured PostGreSQL database. The web interface, running as a Apache Tomcat application, receives queries in a standard XML format, and responds by retrieving real-time or archical sensor data, or meta-data. The same interface is also used to register new sensors, or query the actual capabilities of a sensor, or the sensor network. Each sensor is described by a SensorML file, which is a standard ontology developed for this purpose [2].

[2]http://www.pachube.com
[3]http://www.opengeospatial.org/standards/sos
[4]http://52north.org/

3. MULTIMODAL SENSOR FUSION

The multi-sensor data fusion module collects data from different sensors and modules that process sensory information to produce higher-level information (such as fire and smoke detection results), as well as sensor meta-data. Using a range of fusion techniques, the aim is to reach the final decision about fire events, and determine the level of alarm in each single case.

Data fusion has a very broad scope, and includes methods that fuse information at different levels. When it comes to the accuracy of fusion, we can think of computationally heavy approaches like Bayesian methods and Demster-Shafer methods. On the other hand, power considerations in wireless sensor networks call for lightweight communication and processing requirements for the sensor nodes. In our design, we take these requirements into account.

The sensors in our scenario are distributed on the field, and are affected by natural conditions like wind, rain, and sometimes fire. Temperature and humidity sensors may provide valuable information, but the propagating fire may consume some of the sensors. The smoke detection algorithm may work robustly with a static camera, but wind-induced movements may (and actually do) cause many false positives. These cases illustrate the primary need for data fusion in the FIRESENSE system: To increase robustness under natural conditions. This goal is also related to the issue we raised in the previous paragraph. Reducing the false positives can mean increasing the lifetime of a sensor, as during increased alarm levels, sensors will be queried more often.

In the next two subsections, we describe two different fusion schemes for two different subtasks. In Section 3.1 we discuss fusion of infrared and visible images for improving fire detection via multiple camera systems. In Section 3.2, we describe a fusion scheme to reduce false positives in smoke and fire detection.

3.1 Fusion of infrared and visible images

A fundamental problem in multi-modal image integration is that of aligning images of the same scene (or similar ones) taken by cameras of different modalities. This problem is known as *image registration* and the objective is to recover the correspondences between the images. Once such correspondences have been found, all images can be transformed into the same reference, enabling to augment the information in one image with the information from the others.

In the FIRESENSE project, the infrared (long wavelength) image and the visible image are fused via registration. Both modalities provide different but useful information, and by integrating non-overlapped information, the system can take more accurate decisions. The registration of the Infrared Image (IR) and the Visible Image (ViS) is challenging, because the electromagnetic wavelengths of the ViS sensor and the IR sensor are quite different. Consequently, image properties and patch statistics of corresponding features might be quite different when comparing and matching feature points. Therefore, traditional approaches based on matching/aligning interest points between two images (e.g. the SIFT+RANSAC approach) do not work properly in this application.

Many approaches have been proposed for automatically registering IR and ViS images. Edge/gradient information is one of the most popular features, as their magnitudes and orientations may match between infrared and visible images [15]. In [5], authors first extract edge segments, which are then grouped to form triangles. The transform can be computed by matching triangles from the source to destination images. Huang *et al.* [14] propose a contour-based registration algorithm, which integrates the invariant moments with the orientation function of the contours to establish the correspondences of the contours in the two images. Normally it is difficult to obtain accurate registration by using contour-based methods, because precisely matching all contours detected from two images is challenging. Moreover, this method drastically increases computation time compared to interest point-based registration. As an improvement, Han *et al.* [10] proposed to find correspondences on *moving* contours. An alternative [4] is to make use of moving object paths generated by an object tracking algorithm, as finding correspondences between trajectories helps to align images. This type of algorithms work well when moving objects are available, but our application setting does not assume this precondition.

To deal with the image registration problem, we propose to explicitly align lines derived from edge pixels. While the interest points extracted from both images are not always identical, most major edges detected in one image have correspondences in the other image. Additionally, we focus on aligning the geometric structure formed by a group of lines, instead of descriptor-based individual feature matching, resulting in more accurate feature matching.

Fig. 2 shows an overview of our image fusion approach with an example scene, captured via IR and ViS sensors. The ViS modality is grayscale and histogram-equalized. In the edge detection and line extraction stages, we use a Canny operator to extract edge pixels, followed by a Hough transform to generate straight lines. This transform often produces a bundle of closely positioned lines, which is undesirable. In the next stage, we perform a line fit to determine the best edge representing a bundle. Near-duplicate lines are removed.

To match the lines in two images, we allow each line to have a maximum of three corresponding lines on the second image. Once these correspondences are generated for all lines, we exhaustively check all the geometric structures formed by line quadruplets. A predefined metric helps us to find the best matching. In the last module, we compute a perspective transform matrix based on four corresponding lines.

We have tested our algorithm with six pairs of IR and ViS images/videos, four outdoor and two indoor scenarios, respectively. The minimum length of a line was set to 40 pixels. Our algorithm registered five pairs of images correctly, and failed on one case. There were not enough linear structures for geometric matching in the failure case. We also compared our algorithm with existing algorithms based on interest point matching, all of which performed very poorly. The technical details and quantitative evaluation can be found in [11].

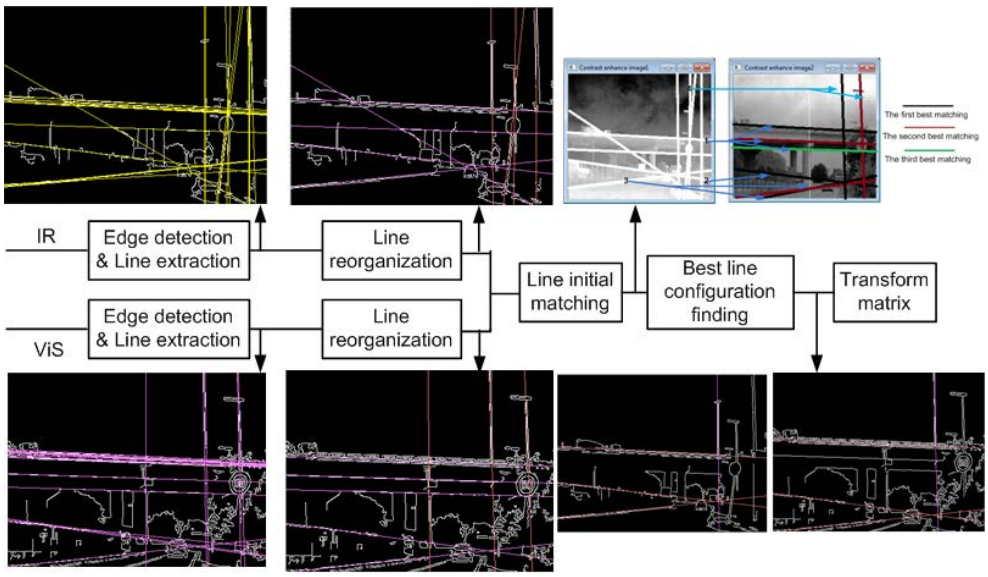

Figure 2: The overview of the image fusion algorithm.

3.2 Fusion of camera and inertial sensors

The wind-induced motion of the camera is one of the major factors contributing to the false positives in the fire and smoke detection module. In [18], it is noted that finding the jittery frames caused by the shaking camera is a major difficulty, if only image-based information is available. For this purpose, Mikolajczyk and Uemura registered video frames using homography [19]. In our application, the camera is significantly far from the viewed objects, and we can assume the entire image to exist on a single plane. Thus, an image based solution would be to look at the mean squared pixel error between consequent frames under minor camera motion perpendicular to the image plane, expressed by $\{\Delta_x, \Delta_y\}$:

$$\arg \min_{\Delta_x, \Delta_y} ||I^{t-1}(x, y), I^t(x + \Delta_x, y + \Delta_y)||, \qquad (1)$$

where $I^t(x, y)$ represents the pixel values of a region of interest of the image at time t.

Such an image based method is able to detect sharp camera movements on the whole, but it has two significant drawbacks. Firstly, this method is very costly, as it scales with the size of the region of interest (which can possibly the entire image, minus some border area removed due to image shifts), as well as the ranges of both Δ_x and Δ_y. Secondly, the acquired images are typically compressed to reduce the computational burden, and small movements of the camera, while significant in producing false positives, can be missed by this method. There are more elaborate schemes in the literature to deal with motion induced problems with image-based approaches [16, 17], and built-in optical image stabilization techniques [20], but these do not apply to our case, as the problem is less in compensating for the blur and more in detecting the problematic frames.

In the virtual reality literature, inertial sensors are employed to position the camera exactly within a given frame of reference [13, 12, 22, 23]. Inertial sensors, when used for precise localization, face the problem of drift, where measurement error accumulates over time. For this reason, hybrid methods are proposed in the literature. For our application, the precise camera location is not relevant. We can use an accelerometer affixed to the camera to measure the magnitude of wind-induced camera motion. The acceleration component due to gravity needs not be taken into account, as long as the sensor's position is fixed with respect to the camera.

By fusing information from the accelerometer with the image information, we can reduce the false positives by discarding frames with movement. For this purpose, we have employed a Waspmote, which is an Arduino-based sensor board. It has an on-chip accelerometer ($\pm 2g$; 1024 LSb/g) that can provide data with 120Hz, which is much more than what we require. The output from this sensor can be quickly processed to produce an indicator of camera motion magnitude in real time.

We have contrasted the image-based approach and the sensor-based approach in Figure 3, for two different situations. In the first situation, the camera is static, and there is no wind. The accelerometer indicates a low level of fluctuation, and the image based approach does not sense any movement. In the second situation, there is a shaking motion with a small amplitude, which can be induced by wind or by holding the camera. For this motion, the accelerometer clearly registers the movements, whereas the image based approach still indicates no movement. For large movements, both approaches will detect the movement.

In order to assess the impact of compression, we have tested the image-based approach for six different compression levels, at 10 or 24 frames per second sampling rate, and for 352×288 and 640×480 image resolution. The detection results were similar, and we concluded that the approach does not visibly benefit from increasing the resolution, frame rate, or compressed image quality.

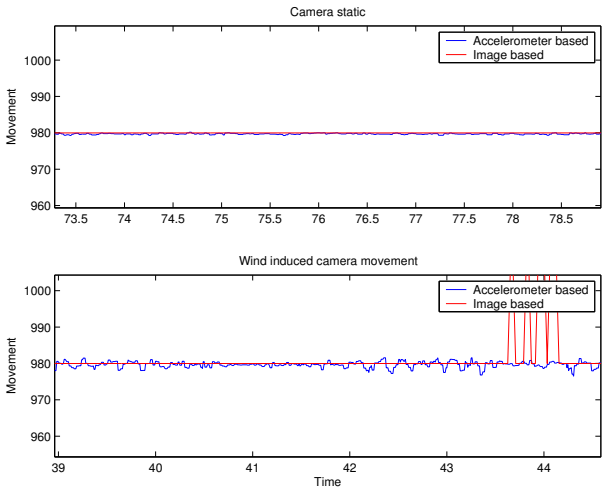

Figure 3: Accelerometer-based and image-based camera movement detection for (top) a static camera, and (bottom) a camera with low-amplitude shaking motion.

4. CONCLUSIONS

The FIRESENSE project aims to design a system for monitoring fires in a real and practical setting. For this reason, it is not enough to design accurate algorithms, but speed, usability, robustness and overall cost are also essential concerns. By employing simple sensors like accelerometers, the speed and cost are kept low. The reduction of false alarms improves the usability of the system. The fusion of two image modalities provides robustness and ensures continuous operation of the system. The proposed methods will be tested and deployed to actual cultural heritage sites, including five pilot sites in Greece, Turkey, Tunisia and Italy.

5. ACKNOWLEDGMENTS

The authors would like to thank Dr. Nicola Marchetti. This work is supported by the FIRESENSE FP7 Project.

6. REFERENCES

[1] L. Bermudez, P. Bogden, E. Bridger, T. Cook, C. Galvarino, G. Creager, D. Forrest, and J. Graybeal. Web feature service (WFS) and sensor observation service (SOS) comparison to publish time series data. In *Int. Symp. on Collaborative Technologies and Systems*, pages 36–43. IEEE, 2009.

[2] M. Botts and A. Robin. Sensor model language (SensorML). *Open Geospatial Consortium Inc., OGC*, pages 07–000, 2007.

[3] S. Calderara, P. Piccinini, and R. Cucchiara. Vision based smoke detection system using image energy and color information. *Machine Vision and Applications*, 22:705–719, 2011.

[4] Y. Caspi, D. Simakov, and M. Irani. Feature-based sequence to sequence matching. *International Journal of Computer Vision*, 68:53–64, 2006.

[5] E. Coiras, J. Santamaria, and C. Miravet. Segment-based registration technique for visual- infrared images. *Optical Engineering*, 39:282–289, 2000.

[6] K. Dimitropoulos, K. Kose, N. Grammalidis, and E. Çetin. Fire detection and 3-D fire propagation estimation for the protection of cultural heritage areas. In *ISPRS Technical Commission VIII Symposium*, 2010.

[7] F. Gomez-Rodriguez, B. Arrue, and A. Ollero. Smoke monitoring and measurement using image processing: application to forest fires. In *Proc. SPIE*, 2003.

[8] N. Grammalidis, E. Çetin, K. Dimitropoulos, F. Tsalakanidou, K. Kose, O. Gunay, B. Gouverneur, D. Torri, E. Kuruoglu, S. Tozzi, A. Benazza, F. Chaabane, B. Kosucu, and C. Ersoy. A multi-sensor network for the protection of cultural heritage. In *Proc. EUSIPCO*, 2011.

[9] Y. Habiboglu, O. Gunay, and E. Çetin. Flame detection method in video using covariance descriptors. In *Proc. ICASSP*, pages 1817–1820, 2011.

[10] J. Han and B. Bhanu. Fusion of color and infrared video for moving human detection. *Pattern Recognition*, 40:1771–1784, 2007.

[11] J. Han, E. Pauwels, and P. de Zeeuw. Visible and infrared image registration employing line-based geometric analysis. *Submitted to MUSCLE Int. Workshop on Computational Intelligence for Multimedia Understanding*, 2011.

[12] Z. Hu, U. Keiichi, H. Lu, and F. Lamosa. Fusion of vision, 3D gyro and GPS for camera dynamic registration. *Pattern Recognition*, 3:351–354, 2004.

[13] Z. Hu and K. Uchimura. Real-time data fusion on tracking camera pose for direct visual guidance. In *Proc. IEEE IVS*, pages 842–847, 2004.

[14] X. Huang and Z. Chen. A wavelet-based multisensor image registration algorithm. In *Proc. ICSP*, pages 773–776, 2002.

[15] Y. Kim, J. Lee, and J. Ra. Image registration based on intensity and edge orientation information. *Pattern Recognition*, 41:3356–3365, 2008.

[16] G. Klein and T. Drummond. A single-frame visual gyroscope. In *Proc. BMVC*, pages 529–538, 2005.

[17] S. Ko, S. Lee, and K. Lee. Digital image stabilizing algorithms based on bit-plane matching. *IEEE Trans. Consumer Electronics*, 44(3):617–622, 1998.

[18] J. Liu, J. Luo, and M. Shah. Recognizing realistic actions from videos "in the wild". In *Proc. IEEE CVPR*, pages 1996–2003, 2009.

[19] K. Mikolajczyk and H. Uemura. Action recognition with motion-appearance vocabulary forest. In *Proc. IEEE CVPR*, 2008.

[20] D. Sachs, S. Nasiri, and D. Goehl. Image stabilization technology overview. *InvenSense Whitepaper*, 2006.

[21] B. Toreyin. Fire detection algorithms using multimodal signal and image analysis. *PhD thesis, Bilkent University, Department of Electrical and Electronics Engineering, Ankara, Turkey*, 2009.

[22] S. You, U. Neumann, and R. Azuma. Hybrid inertial and vision tracking for augmented reality registration. In *Proc. IEEE Virtual Reality*, pages 260–267, 1999.

[23] S. You, U. Neumann, and R. Azuma. Orientation tracking for outdoor augmented reality registration. *IEEE Computer Graphics and Applications*, pages 36–42, 1999.