Image analysis platform for data management in the meteorological domain

Igor G. Olaizola, Naiara Aginako, Mikel Labayen
Vicomtech Applied Research Centre
Department of Digital TV & Multimedia Services
Mikeletegi Pasealekua, 20009 San Sebastian, Spain
iolaizola@vicomtech.org, naginako@vicomtech.org, mlabayen@vicomtech.org

Abstract

This paper proposes an architecture to provide semantic media information to the current existing meteorological models and prediction techniques. Satellite images have been used by meteorologists during the last 50 years, but we present a new method to take advantage of local images taken from the earth. Networked terrestrial weather stations can offer valuable image information, both of local and wide areas adding details that cannot be captured by satellites. Based on the results of two projects carried out together with the Basque Meteorology Agency (Euskalmet), we propose a method to port from image data to semantic meteorological information and an architecture to integrate the existing weather data and knowledge estructures with multimedia semantics. The validation of the analysis system has been carried out using sky images taken in visual spectrum and the results have demonstrated the great potential of such platforms that could be extended to other data sources in order to apply multimedia semantic technologies in application fields like meteorology.

1. Introduction

The last 100 years science of meteorology have been based on the numerical analysis of large amounts of data collected from the environment [10]. During the last 50 years (with the launch of *Vanguard II*, the first weather satellite[9]) apart from numerical data like temperature, pressure, humidity, wind, etc., multispectral images have improved dramatically the analysis and prediction models of the meteorological behavior[13, 6] including water vapour images, visible images, IR images, microwave images, . . .

Nowadays, satellite image acquisition and processing techniques's offer precise information about environmental features in very wide areas. This fact has allowed a deep improvement in the knowledge regarding to meteorology/climatology behavioral models and a much higher accuracy in weather forecasts.

However, global information has to be complemented with local data when the region of interest is limited to a specific geographic zone. This local data is provided by terrestrial weather stations equipped with classical measurements instruments (termometers, higrometers, pluviometers, barometers, etc.). Some other information like the local degree of cloudiness has to be provided by human observers. For example, cloudiness degree is normally given in octets, and the precision depends on the observer's skill. In this context, the current state of the art in image acquisition and processing and existing wireless data transmission technologies add a new set of possibilities at an affordable cost. Numerical information obtained by measurement instruments can be extended with *visual concepts* and analyzed by combining data-mining and ontologies. It increases the accuracy of results and helps to automatize the entire data extraction process. Moreover, networked local information processing units can provide global information, improving the data collected by satellites.

The remaining of this paper is organized as follows. Section 2 shows the link between the local semantically enriched data processing and global meteorological/climatological models and platforms. The semantic processing of the concepts resulting from the image analysis platform will be presented in Section 3. In Section 4 we describe the implemented image acquisition/processing platform where meteorological concepts are obtained. In addition, experimental validation (precision and recall) of each class segmentation is presented. Finally, some conclusions are presented in Section 5.

2. From multimedia semantics to weather knowledge

The science of meteorology combines a physical models (where the meteorological behavior is explained mathematically) which could be considered as explicit knowledge, and knowledge management techniques (based on data mining / machine learning techniques or ontologies). Explicit knowledge can be introduced in the system by ontologies that can then infer new relationships among concepts. On the other hand, implicit knowledge (knowledge that has not been discovered or that cannot be described in a straightforward manner) can be extracted by using data mining algorithms. These to methods may be interdependent in order to improve the results of knowledge extraction and it is specially useful for ontologies [11]. Figure 1 shows a simple diagram of the knowledge acquisition flow, where the first steps are oriented to filter the information and to extract the useful one. Then knowledge extraction processes can be applied.

When the data that meteorologists are going to deal with is highly complex (radar information, multispectral images, etc.), a previous work has to be done to extract useful information from them. Feature selection is one of the key factors to obtain meaningful information. For example, in the *Skeye* project, color related features and entropy (related with region's texture) were selected as most significant characteristics for object segmentation. The presented image segmentation process (Section 4) and the framework shown in Figure 4 establish the basis to automatize part of the work that human observers have to do manually. The potential of this approach is not limited to the fact that some work is automatized, but it allows the integration of many networked weather stations that can cover big areas[2]. This integration can provide information to improve the existing meteorological models or to create new behavioral patterns focused on local weather particularities. Thus, in the same way that numerical parameters are used to build mathematical models, the concepts extracted from image analysis contribute in the meteorological knowledge domain. This is particularly useful to find new relationships among apparently independent phenomena (e.g: the long-term incursion of warm water in the Peruvian coast known as "El Niño" [1] has strong climatic consequences in some other parts of the world like Africa and Southeast Asia).

3. Semantic representation and processing

Sky images contain a lot of semantic information that can be extracted using image processing and data mining techniques. The *Skeye* project [7] is focused on the extraction of classes that are normally estimated by human observers (cloudiness, fog, rain,...). Apart from the main classes, these are not the only concepts that can be detected and handled by the semantic processing module. New classes like *snow*, *hail*, *dew*,etc. can add knowledge about the weather situation at a specific moment and if this information is combined with other stations, high detail global maps could be created.

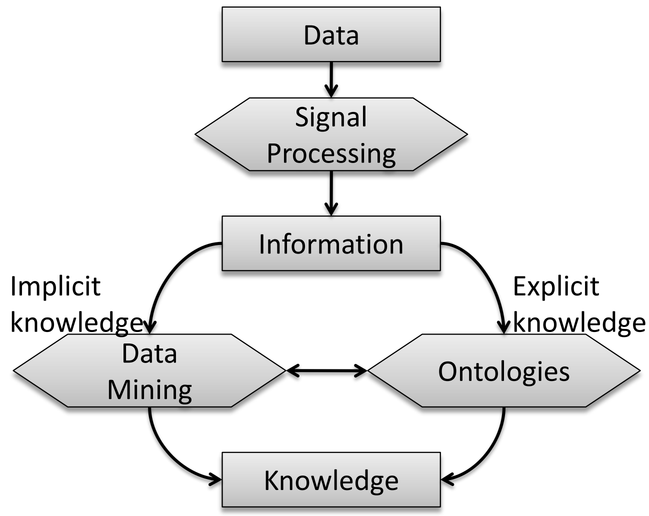


Figure 1: Knowledge acquisition process

There are several initiatives to define a common way to represent geographically referenced heterogeneous data [4, 12](mostly meteorological and environmental information) where ontologies are used to describe the existing explicit knowledge on these topics [8] but still there is not a widely accepted common standard to represent, manage and exchange environmental and meteorological [14] information combining global and local data sources.

3.1. System architecture

Our proposed system is based on the *Skeye* [1] and *ISD* [2] projects, where infrastructures owned by *Euskalmet* [5], (networked meteorological measurement instruments, radar information, etc.) are combined with sky images taken from the earth.

Figure 2 shows the general architecture of the system based on the results of *Skeye* and *ISD* projects.

Data acquisition systems located on remote places are connected to the central unit through wireless communication systems. In the case of the Basque Country (weather stations are located few kilometers away from the central unit or from other weather station, and the orography is very irregular formed by isolated valleys and mountains that ar not too high), WIMAX has been adopted as most adequate technology which offers good bandwidth and reliability at an affordable cost. However, deployments in bigger countries with different orography conditions will need a specific study.

[1] *Skeye is a R&D project funded by the Basque Government and carried out by Euskalment, Dominion and Vicomtech*

[2] *ISD* is a R&D project within *Etortek*, Basque Estrategic Research Program

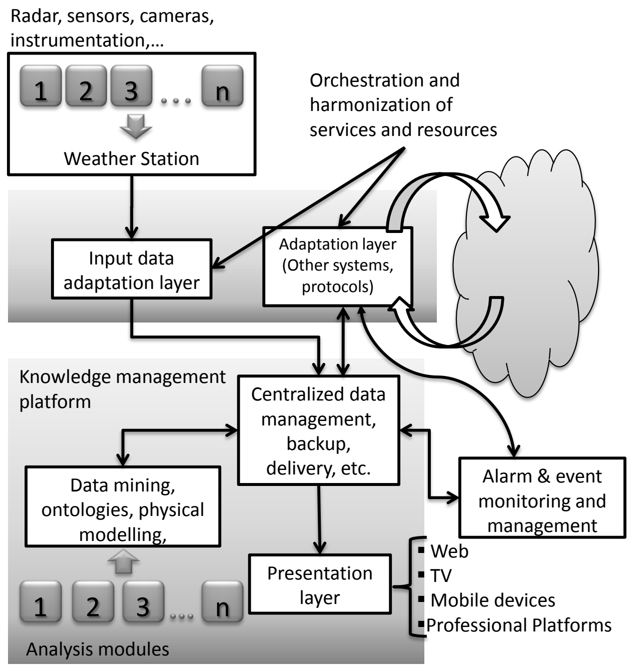


Figure 2: General architecture of the meteorological information management system.

The adaptation layer allows the communication among diverse networks and protocols providing interfaces for interoperability with proprietary devices and external meteorological or risk management systems.

The knowledge platform, is the core of the whole system. This platform is the basis for typical analysis activities of meteorologists and moreover is where machine learning processes are carried out. The knowledge platform can interchange data with different external data analysis modules (mathematical tools, data mining libraries, ontologies,...) and results can be shown in a virtual georeferenced map (Figure 3).

3.2. Ontologies and data mining

The huge amount of information centralized in the *knowledge management platform* requires a combination of tools and techniques to exploit the potential of the explicit and implicit knowledge. A combination of data-mining algorithms and ontologies proposed in this paper (Figure 4) is based on the model described by Bogorny et al. [3] and the cycle of data mining with ontologies presented by Nigro et al. [11]. It can be observed that data mining algorithms are used as first step to filter the data and reduce the dimensionality and to make it affordable by ontologies. Moreover, data mining processes can extract the implicit knowledge and hidden relationships among different data sources. A second ontology is used to guide the data flows and adjust the data mining algorithms. This second ontology improves the efficiency of the data mining algorithms by adding heuristics to the process.

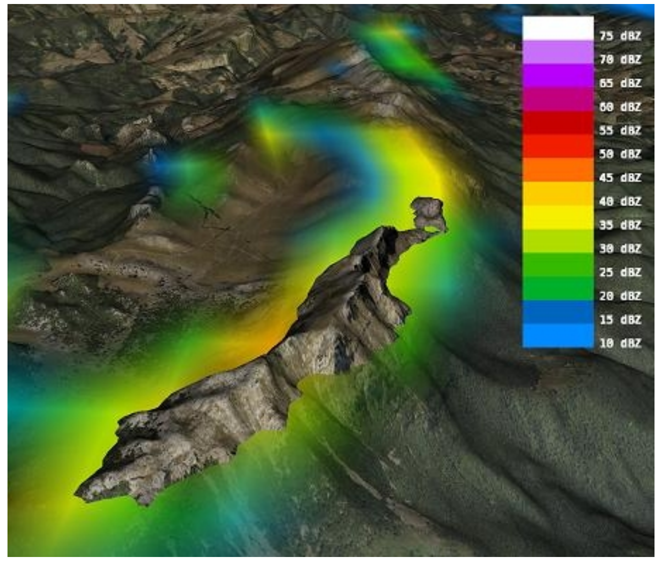


Figure 3: Virtual representation of suspended water (clouds and rain) on a georeferenced map

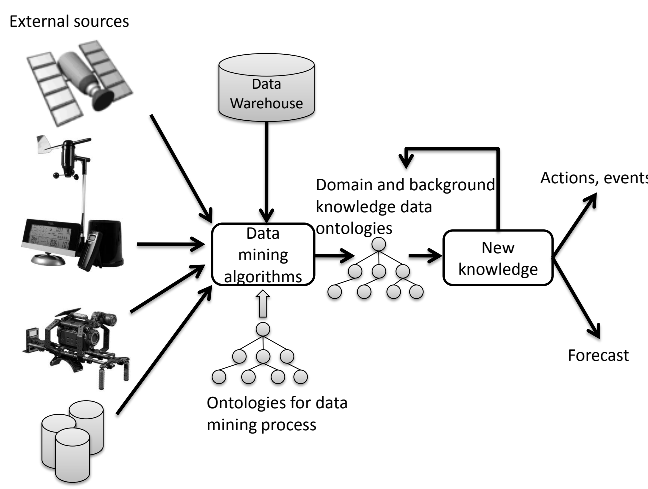


Figure 4: Framework of data mining with ontologies.

4. Meteorological image processing platform

The image acquisition and processing part (where different objects are segmented) presented in the previous Section 3 (Figure 2) has been reflected in *Skeye* [7], with which low-level features have been validated. While the mentioned paper explains the internal details of segmentation algorithms, this paper shows *Skeye* at system level, describing how it is integrated within a general meteorological analysis platform.

The images are processed using digital image processing techniques described in [7] in order to obtain local meteorological concepts based on five basic classes segmentation: *earth, sky, sun, clouds* and *fog*. This basic concepts mainly provide additional data (cloudiness or fog detection) regarding the local weather conditions.

The Figure 5 shows the general description of the system.

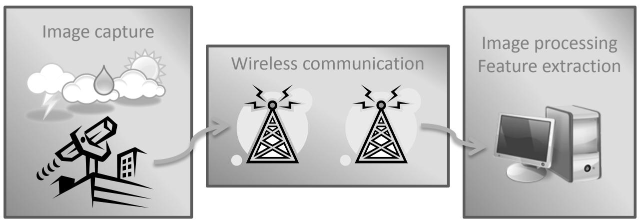


Figure 5: Skeye block diagram [7].

4.1. Modules

This section summarizes the functionality of each module which conform the system, as well as their specifications:

Image capture A camera located at terrestrial meteorology station takes pictures of the sky covering the complete celestial dome.

The quality and properties of the images, which will condition the segmentation accuracy, depend on how the camera's shutter speed, diaphragm aperture and white balance are set.

Wireless communication The retrieval of the information provided by all remote meteorological stations is made through WIMAX (IEEE 802.16) infrastructures.

Image processing This module, explained in the next section, centralizes all the information coming from the WIMAX network, it analyzes the images and it extracts features using digital image processing techniques in order to segment the image in five classes.

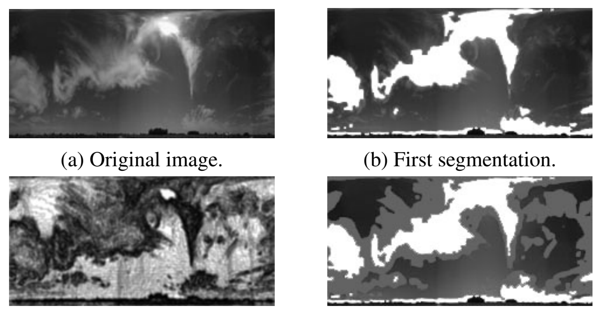
4.2. Implementation of the Image Processing module

The most outstanding functionality of *Skeye* is the capability to segment the image in five predefined classes: *earth, sky, sun, clouds* and *fog*.

Firstly, the earth class is segmented. The histogram of B channel in RGB color space represents two easily separable pixel densities using dynamic threshold. They represent the group of pixels belonging to the earth and the rest of the classes.

Secondly, *Skeye* separates the non-covered sky pixels, characterized by a chromatic component, for the sun and clouds, characterized by an achromatic component. Nevertheless, the non-covered sky can tend to be achromatic in some areas, as well as some transparent clouds reveal the blue hue of sky.

In order to resolve this problem, the saturation component is binarized (HVS color space) with conservative threshold extracting the first non-covered sky pixels. After that, the entropy values of the co-occurrence matrix are used in textural feature extraction to separate definitively the rest of pixels belonging to non-covered sky class.



(a) Original image. (b) First segmentation.

(c) Entropy values. (d) Second segmentation.

Figure 6: Non-covered sky segmentation [7].

The intensity values of sun class pixels are represented with the maximum value of the grey level histogram as well as some clouds pixels if those are near to sun. The constant area, characteristic circle shape and orientation properties of the sun region are used for the final segmentation.

Finally, the fog detection is tried inside the cloud class. If the clouds cover parts of the earth, which is usually visible at sunny day, they consider part of fog class. The applied techniques are two and their based on the earth class constant shape when it has been captured without fog: detection based on pattern recognition and detection based on earth silhouette shape analysis.

4.3. Experimental Validation

For the experimental validation of the concept extraction module, we used a database provided by *Euskalmet*, the Meteorology Agency of the Basque Country[5]. These test images were taken at different time during daytime (from 11:00 to 15:00, 2 images/2 hours) in autumn season and under diverse meteorology conditions.

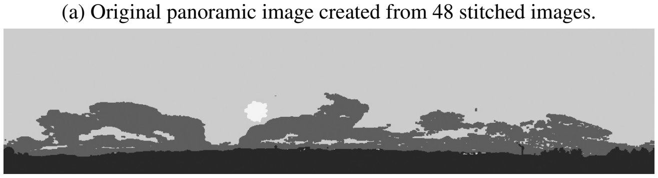
We carried out the evaluation using 780 images of the database. For the experiment we divided the image collection in two groups depending on whether they contain earth class or not. *Sky, Sun* and *Clouds* classes extraction is validated in the entire database while *Earth* and *Fog* classes extraction is validated only in the 480 images containing earth class. Table 1 shows the precision and recall obtained in each class segmentation.

Classes	Precision	Recall
Sky	0.92	0.68
Earth	0.83	0.54
Sun	0.67	0.85
Clouds	0.89	0.63
Fog	0.81	0.75

Table 1: Segmentation results, precision and recall.

An example of the successful segmentation is shown in Figure 7.

(a) Original panoramic image created from 48 stitched images.



(b) Segmented image.

Figure 7: Segmentation result.

During the validation of the content extraction module, we identified several limitations due to the use of cameras operating in the visual spectrum. These kind of cameras don't require big investments and thus can be deployed along big weather station networks, but their high dependency on light conditions needs the above explained post-processing task to extract the wanted information and obviously, they cannot be used at night.

Instead, **infrared cameras**, offer higher SNR in most conditions where visual spectrum cameras can be used and under extreme conditions they still keep this high contrast. The so called "Thermal" cameras can detect the radiation emitted by clouds and therefore there are not limitations due to the daylight. On the other hand, the region near the sun is always too bright to be directly captured in visual spectrum, but with infrared cameras[3] are not sensitive to this brightness as can be seen in Figure 8. It means that the same IR camera can be used under different extreme condition without any changes in the set-up.

[3] They may be keep protected from long sun exposures to avoid breakdowns

Although, the better performance of IR cameras simplifies the content extraction process, the high investment needed for the installation of an network equipped with such cameras makes this approach not viable for most real deployments. This fact remarks the necessity of the above described technique for visual image segmentation and content extraction.

Figure 8: Sky picture looking directly at the sun, taken with an IR camera (spectral range of the camera: $7.5 - 13.5\mu m$).

5. Conclusions

A new application domain for multimedia semantics has been proposed in this paper. The common use of semantic technologies in scientific fields like meteorology and in multimedia content management offers a new set of possibilities to improve the modeling and forecasting processes.

The promising results of the concept extraction module, encourage us to continue with the implementation of a system that merges not only numerical data obtained from different meteorological stations but also visual information. Even more, the system permits to combine this information with other meteorological infrastructures. These results have demonstrated that a closed and dimensionally affordable multimedia domain can be built with good performance results improving both, the quantity and quality of the information available for meteorological models and weather forecasts.

Moreover, the potential of such systems is not limited only to the meteorological domain. Environmental control and monitoring systems will also get a big benefit improving the natural disaster prevention and coordination activities.

6. Aknowledgements

The work described in this paper has been funded by the Basque Governments *ETORTEK* and *INTEK* research and development programs. Moreover, the authors would like to thank *Euskalmet* and Dominion [4] for providing the meteorology background and resources needed to carry out this work.

References

[1] El niño impact. Availabe at: *http://www.oceanmotion.org/html/impact/el-nino.htm*.

[2] Álvaro Segura, A. Moreno, I. García, N. Aginako, M. Labayen, J. Posada, J. A. Aranda, and R. G. D. Andoin. Visual processing of geographic and environmental information in the basque country: Two case studies. In *Geographical Information Processing and Visual Analytics for Environmental Security*, Trento, Italy, October 2008.

[3] V. Bogorny, B. Kuijpers, and L. O. Alvares. Reducing uninteresting spatial association rules in geographic databases using background knowledge: a summary of results. *Int. J. Geogr. Inf. Sci.*, 22(4):361–386, 2008.

[4] S. S. Durbha and R. L. King. Knowledge mining in earth observation data archives: a domain ontology perspective. In *Proc. IEEE International Geoscience and Remote Sensing Symposium IGARSS '04*, volume 1, Sept. 20–24, 2004.

[5] Euskalmet. Basque Meteorology Agency. Available at: *http://www.euskalmet.euskadi.net*.

[6] A. K. Maini and V. Agrawal. *Satellite technology, principles and applications*. John Wiley & Sons, Ltd., 2007.

[7] M. Labayen, N. Aginako, and I. García. Weather analysis system based on sky images taken from the earth. In *Proceedings of VIE 2008 - The fifth International Conference on Visual Information Engineering*, pages 146–151, Xian, China, July 2008. IET, Michael Faraday.

[8] C. G. N. Macário and C. B. Medeiros. Specification of a framework for semantic annotation of geospatial data on the web. *SIGSPATIAL Special*, 1(1):27–32, 2009.

[9] C. McLaughlin Green and LomaskMilton. *Vanguard, a history:*. The NASA Historical Series, Washington, DC, 1970.

[10] F. Nebeker. *Calculating the weather: metheorology in the 20th century*, volume 60 of *International geophysics series*. Academic Press Limited, 1995.

[11] H. O. Nigro, S. E. González Císaro, and D. H. Xodo. *Data Mining with Ontologies: Implementations, Findings, and Frameworks*.

[12] P. S. Pradhan, R. L. King, T. Haupt, and V. Anantharaj. The cyberinfrastructure and image information mining. In *Proc. IEEE International Geoscience and Remote Sensing Symposium IGARSS '04*, volume 1, Sept. 20–24, 2004.

[13] P. Santurette and C. Georgiev. *Weather Analysis and Forecasting: Applying Satellite Water Vapor Imagery and Potential Vorticity Analysis*. Elsevier Academic Press, 2005.

[14] WMO. World Meteorological Oganization. Available at: *http://www.wmo.int/pages/index_en.html*.

[4] `http://www.dominion.es`