Contextual Geospatial Picture Understanding, Management and Visualization

Stavros Christodoulakis
Lab. of Distributed Multimedia Information Systems and Applications
Technical University of Crete
stavros@ced.tuc.gr

Michalis Foukarakis
Lab. of Distributed Multimedia Information Systems and Applications
Technical University of Crete
foukas@ced.tuc.gr

Chrisa Tsinaraki
Lab. of Distributed Multimedia Information Systems and Applications
Technical University of Crete
chrisa@ced.tuc.gr

Eleni Kanellidi
Laboratory of Distributed Multimedia Information Systems and Applications
Technical University of Crete
ekanellidi@gmail.com

Lemonia Ragia
Department of Architecture
Technical University of Crete
lemonia.ragia@isc.tuc.gr

ABSTRACT
We describe *SPIM (Spatial Image Management)*, a framework and a system that allow identification and visualization of the semantics of distant objects in nature, which exist in pictures taken by mobile devices. Semantic objects in nature are represented on 3D maps and tagged with semantic and spatial descriptors. The system functionality creates, on top of pictures, a layer that allows semantic user browsing in augmented geospatial images, semantic object identification in pictures, object interaction, contextual map visualization, and interlinking of information in digital libraries and the web. We present requirements of semantic geospatial descriptions, system functionality, matching algorithms and experimental results.

Categories and Subject Descriptors
H.3.7 [**Information Storage and Retrieval**]: Digital Libraries – collection, dissemination, standards.

General Terms
Algorithms, Experimentation.

Keywords
Mobile Context, Semantic Geotags, Semantic Maps, Ontology, GeoSpatial, Multimedia Digital Library, Mobile Camera, GPS, Compass, Mobile Sensors, Tourism, Culture.

1. INTRODUCTION
Mobile user context interrelated with geospatial semantic content can enhance, by supporting very important multimedia functionalities, applications like travel, culture, tourism, and multimedia digital libraries. In this paper, the user context includes the user interests at a certain time point, as well as the spatial context captured by sensor devices (GPS, compass, etc.) at the time of picture taking from a mobile device such as a mobile phone or sensor-equipped digital camera. Geospatial semantic content includes information and semantics of objects located on the earth surface. They could be nature objects like lakes, or other objects of interest like forts.

In this paper, the application functionalities of interest center around these questions: *(a)* In a picture taken by a mobile device at a certain location, which are the objects of interest to the user, and where are they located on the picture? *(b)* Which are these objects, what are their semantics, what other information is known about them in a digital library or on the web? For large objects nearby, like buildings in a square of a city, such questions may be relatively easy to answer using the GPS and compass sensors and the topology of the square. These questions may be very difficult to answer in a generic geospatial setting, since objects at distance may be barely visible to the human eye, the sensor capturing inaccuracies prohibit locating the position of small objects in the picture, and image processing often fails to locate small objects in this context.

In this paper we present *SPIM (Spatial Image Management)*, an integrated framework and a system for managing the semantic spatial information and exploiting it together with contextual information in order to provide image understanding, interaction with image objects, and visualization in multimedia applications.

The rest of the paper is structured as follows: Our model for the semantic description of real world objects and the concept of semantic maps are presented in Section 2 and our approach regarding the use of image descriptors is discussed in Section 3. The contextual parameters that we use and their exploitation are described in Section 4 and the picture content prediction approach of the SPIM framework is presented in Section 5. The system functionality is described in Section 6. Image registration and semantic spatial object identification are discussed in Section 7, along with the experimental evaluation results. The paper concludes in Section 8, where we also discuss our future research directions.

2. SEMANTIC DESCRIPTION OF REAL-WORLD OBJECTS – SEMANTIC MAPS
We use a model of *semantic entities* of the real world to describe objects and object types. We assume that objects of interest to users are described by *domain ontologies*. For example, a user interested in "Greek Archaeology" will be using the ontology in this domain and will be taking images of objects that correspond to these concepts. An *individual* is a real world object that is described by an ontology concept. A specific temple of Zeus in a

MoMM2013, 2-4 December, 2013, Vienna, Austria.

given location is an individual. An ontology is useful, because it provides a consistent way to refer and describe semantic concepts. It also provides a means for semantic interoperability between the user and the system or among users that have common interests [1][7].

Traditional digital maps usually contain geographical information or just names of populated places, have a defined scale and show a (usually) rectangular area as seen vertically from an elevated position. *Semantic maps* on the other hand may be 2D or 3D and may have many additional capabilities [2] [5].

The first important characteristic of semantic maps is that they have the functionality to associate each pixel on the map with geographic coordinates. This enables them to represent any kind of spatial information taken from any angle and not just from above with the top side of the map pointing north. In addition, any image can be used as a semantic map if processed in that way, for example old maps or archaeological drawings can be digitized and processed as semantic maps.

The second important characteristic is that they enable the integration and visualization of semantic objects overlaid on top of them. The semantic objects have geographic representations and they can be visualized either by transforming them to pixel coordinates or by enforcing a different pixel based (and probably more accurate) representation in advance. The semantic objects are not only shown on top of the map, they can be interacted with and all the information that has been associated with them becomes visible. Semantic maps can additionally visualize geo-referenced photos (with location as well as direction information), which may also be interactive.

A semantic map, in addition to the spatial information about geographic concepts, contains also the footprints (polygons consisting of GPS points) of semantic objects represented on the map. Those objects may be instances of object types described in an ontology. For example, a given footprint on a map may represent the Byzantine church of "Santa Maria in Samaria", which is an individual of the type "Byzantine Church" of the ontology "Greek Archeology". Individuals from different ontologies may be represented on the same map. A geographic ontology (with types that include cities, villages, mountains, rivers, sea, etc.) is represented in a similar manner.

The semantic model incorporates domain ontologies recursively composed of semantic types with semantic attributes, which have individuals. Semantic maps also have a type and semantic individuals. The semantic information is separated from its map representation. The same semantic information can be represented and visualized in different 2D or 3D semantic maps.

The system can generate personalized versions of semantic maps by visualizing only footprints of the semantic objects that are of interest to the user. By selecting the ontologies of interest along with the semantic types of interest or even single individuals, the user is able to fully customize the generated semantic map. For example, the user may want to see a map with only the footprints of all the Byzantine churches in the area, or a map with the "Church of St. John". Therefore, personalized semantic maps contain fewer and more relevant semantic entities, which is desirable for small screens. Personalized printing services can also be offered with this functionality.

3. IMAGE CONTENT DESCRIPTIONS

In the simplest case, image content descriptions allow image indexing for retrieval [7][8]; they provide functionality essentially similar to that found in multimedia Digital Libraries, including image unique identification, image and segment management, association of semantic concepts and individuals with images and segments, filing and interlinking, querying and browsing, etc.

Images taken by a user are identified by a unique ID. One reason is that some parameters of the image (like GPS position, direction, etc.) may have to be extracted by different devices, recorded and transferred to be integrated in the remaining image descriptions. Associating the contextual metadata of an image with the semantic maps allows us to predict the semantic objects described by the ontologies and individuals of a semantic map that are within the image. The algorithm to do that selects the footprints of the semantic objects that are included or overlap with the visible by the camera landscape. Images may be linked to provide guided tours through an image collection.

Images may have specific regions identified within them. The identification of the regions can be done with a segmentation graphics package, or automatically with segmentation software. A region may be composed of more than one geometric segment. Regions may be used to show in an image where real world semantic individuals are presented. For example, regions may show where a mountain, a fort or a village is presented. We associate real world individuals with images or image regions. We can then answer questions on what images contain within them and vice versa, given a certain region of an image, which are the semantic entities that are there, and in which other images and regions the same semantic entities are shown.

An important consequence of this scheme is that it can automate a lot of work. If there are many different images that show the same set of semantic objects, the semantics of the entities of the real world do not have to be described several times in a time consuming manner. We only need to associate the IDs of the regions with the IDs of the corresponding individuals. This may be done automatically if certain image parameters are known.

4. CONTEXTUAL PARAMETERS

The information contained in the Exif [4] image metadata is useful for automating image content description and associating digital image contents with spatial content. We use image taking parameters such as the focal length, device parameters such as image sensor size, GPS parameters such as location and altitude and direction parameters such as angle. By combining these parameters, we can specify the position and direction of the image, as well as the angle of view of the camera. In some cases, compass direction and camera tilt information is not supported by the mobile device. The system will have to explore alternative algorithms to compensate for it.

The Exif metadata also contain a field called "subject distance", which, if known, can be used for the calculation of the minimum distance after which and the maximum distance before which objects are in focus. Unfortunately not all the cameras in the market today provide this value. When range finding devices are not available, we can still compute an approximation of the subject distance using two or more images from different locations. This can automate the image processing later on. When the "subject distance" is known or calculated as we

described above, the image content descriptions can be modified to show only those objects that are in focus.

Note that we utilize the Exif metadata for two reasons: *(a)* They are automatically provided by the digital cameras; and *(b)* The metadata description capabilities of Exif are a subset of the MPEG-7 [6] capabilities that we have used in our previous work for the semantic description of multimedia content [9].

5. PICTURE CONTENT PREDICTION

In the remaining sections we consider 3D semantic maps for the sake of simplicity. When taking an outdoors photograph, not everything between the user and the direction of image taking is visible. There may be objects blocking the view. For example, hills may hide objects in valleys. The algorithm that calculates the visible objects from a certain camera position and direction traces rays that begin from the camera, follow the direction of the camera and are within the angle of view. At regular intervals, it checks if the current GPS position is visible from the point of view and if a semantic object is visible at this position. To do that, it has access to both elevation data and semantic maps enriched with semantic objects. The semantic objects in the semantic map knowledge base are represented by GPS coordinates, thus their footprints can be visualized on any semantic map of the area they are situated. Therefore, the semantic information is separated from its map representation.

As a result of associating the image with the spatial information, we can for example automatically predict the geographic objects (mountains, cities, lakes, etc.) that appear in an image at the time of image processing, using the contextual parameters of the picture capturing; this way, we can extract and store spatial metadata about the image. We can also do it in real time, on demand, utilizing the parameters calculated, the spatial model and the maps of the area. This could be useful for the user orientation and tourism applications.

6. SYSTEM FUNCTIONALITY

We have designed and implemented components for managing digital images with spatial content and integrated contextual metadata. SPIM incorporates semantic maps and ontologies as described above. A *Semantic Map Server* has been implemented which offers services for creating personalized semantic maps. Information about semantic map aspects is offered as services, or a whole new personalized semantic map is created and transferred to the user site. The personalized semantic map can then be used for digital image metadata creation, for the automatic construction of a visual spatial interface for browsing the contents of the digital image library, as well as exploring the contents of the digital images.

A SPIM component is used for capturing semantic metadata from the digital image and semantic information captured by the sensor devices. This process is taking place when the images are downloaded from the camera into the user PC. The parameters of the model are calculated as described above, and the digital image is associated with the spatial information in the personalized semantic map that the user selects. As result, spatial and semantic metadata about the image contents are generated, and associated with the picture.

The personalized semantic maps are also used as personalized interfaces for visual browsing of the digital library contents. The interface allows the user to generate them on demand by selecting different versions of personalized maps. On top of those maps, the system can display interactively views of the digital library contents. For example, the user can interactively specify that he/she wants to see the footprints of the images that include a specific semantic object (or a set of semantic objects). The system will display the footprints of the images on top of the semantic map. The footprint of an image shows on top of the map the location from which the image was taken, the area covered by the image, etc.

SPIM also offers capabilities for precise image registration so that images can be superimposed with a graphical layer that allows interaction with the semantic objects visible in the picture. This is essentially an expandable augmented reality environment, where the pictures taken allow interaction with semantic objects visible on pictures, which can lead to other pictures, related documentation (archives, history, legal documents and progress in archaeological excavations) or information in public knowledge bases such as Wikipedia.

7. IMAGE REGISTRATION, SEMANTIC SPATIAL OBJECT IDENTIFICATION

The problem of automatically and precisely identifying semantic individuals in a picture (such as a specific village or fort) is difficult when those objects are far. Note that unlike near object images (such as those taken in a square or a road of a city), images of the natural environment may have several difficulties, like contextual parameters capturing errors that become significant for small object identification at distance (errors in GPS, in compass direction, etc.), new objects that do not appear on the geospatial map appearing in the picture (clouds, shadows, etc.), distant objects not visible well depending on the weather (humidity), etc. Note also that mobile devices may only have a subset of the contextual capturing capabilities (having for example only GPS positioning but not compass direction or tilt). These factors suggest that algorithms are needed for precise image registration with view from the specific point.

SPIM tries to match a picture taken with the predicted view of significant (for automatic identification purposes) objects in the direction of viewing (like mountain tops). We note here that semantic objects of nature (like mountain tops, rivers, sea, etc.) are useful to be included in the semantic maps not only for satisfying user curiosity and tourism applications, but also for correct image registration since with their size and distinctness (in colour, etc.) they provide support for correct image registration that would not be possible with smaller objects that may be of interest to the user.

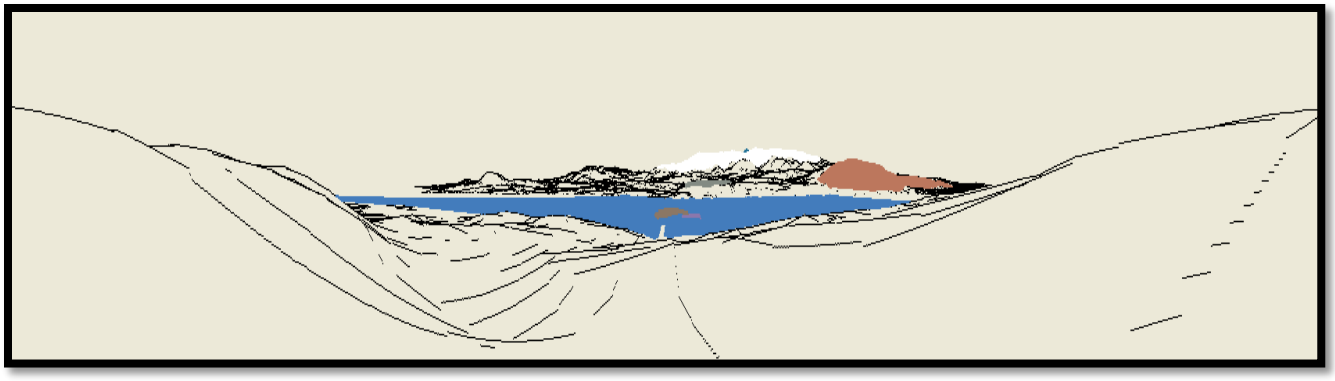

Figure 1. Creation of the panoramic view from a given GPS point using 3D maps. The camera will only view a part of the panoramic view that depends on the camera parameters (near the small island in the following figures).

The percentage of the panorama captured by the picture is calculated first from the camera parameters. Then SPIM segments the image and tries to match the boundaries of certain segments with graphic objects of the predicted view that is visible from the observation point (the part of the panorama calculated). An example is presented in Figure 1, where we describe the creation of the panoramic view from a given GPS point using 3D maps.

The matching algorithm approximates the extracted boundaries from both the photo and the predicted view (2D view model) with line segments. Then it tries to find the optimal match between the approximated line segments. The optimal match provides the best fit transformation for transforming the 2D model to fit the photo. The algorithm is called several times with slightly different parameters and the best result is kept. The matching method is a modified version (that fits our purposes) of the "Steepest Descent Local Search" [3], which can be used for matching any 2D object.

7.1 Experimental Evaluation

Experimentation took place on images from different geographic locations, taken in different times of the year. The dataset we used consists of 31 nature images. All the 31 images contain mountains and land areas right in front of the user, 29 images contain sea areas, but only 6 of these images contain sea areas enlightened from the sun; finally, 3 images contain islands. The goal in each experiment was to evaluate the quality of the resulting augmented image (which is superimposed with a graphical layer and the locations of semantic objects on it – see Figure 2).

The skyline (separation of mountains with sky) is often significant for nature image registration purposes (especially in clear weather and in reasonable distances). In cloudy weather or humid seasons (summer) however, the visibility at distance is not high, and there may be significant errors in the registration process.

We applied our skyline detection algorithm on the evaluation dataset and we ended with 18 satisfactory skyline detections and 13 complete failures (due to low visibility of the distant mountains). Overall, our skyline detection technique worked for 58% of the images. Thus, additional clues are needed for registration purposes.

Another range of algorithms that we have experimented with, exploit the coastline separation of the land with the sea. The algorithm exploits the colour differences between the sea and the shore. This class of algorithms allows correct image registration even with a cloudy sky (see Figure 2 as an example). To apply the algorithm, the semantic map must have descriptions of sea segments as semantic objects. In the case that there is a near and a far coastline the near coastline was found to be more helpful. Problems however arise from the reflection of the sun which breaks the sea into segments. An improvement of the coastline separation algorithms was observed with algorithms that we developed for taking into account the relative location of the sun and its reflection to the sea (taking into account the time that the picture was taken and the date which allows us to calculate the relative location of the sun). Coastline matching may have advantages over skyline matching when high humidity or cloud conditions exist. The experimental results in those conditions showed that coastal lines can be used to improve the results obtained with the skyline detection, or in difficult weather conditions to be used instead of them. Coastline matching may have a disadvantage when the coastline is horizontal in the picture (because it does not allow good scaling). Experimentation with small islands at a distance failed to improve consistently the registration results.

Figure 2. With a cloudy sky the skyline cannot lead to successful image registration. The coastline was successfully used. After registration, the location of the archeological site on the hill on the other side of the bay is accurately pointed.

We applied our coastline detection algorithm on the evaluation dataset and the results were in-line with our expectations: From the 31 dataset images, 26 images had satisfactory coastline detection and we had 5 complete failures, due to the fact that the land area in front of the user contained gray stones.

Finally, we examined the case of inexpensive cameras, where only GPS information is available but not the camera direction. In this case, the panoramic view (360 degrees view of the landscape around the GPS point) is computed, and the image matching is tested with successively overlapping parts of the panoramic view in order to find the best match. Although costly, experimentation showed successful results. For this experiment, we used a different dataset, consisting from 11 images from various places around the area of Chania, which were captured with a simple digital camera with no direction support.

The resulting augmented images were classified into 7 categories according to their quality. From the best to the worst they were: *Perfect*, *Good*, *Acceptable*, *Average*, *Bad*, *Awful* and *Failed*. The first four categories were considered "pass" grades, while the other three "fail" grades. Every resulting augmented image was inspected thoroughly in terms of how well the individual footprints were matched in the picture contents.

The evaluation results are shown in Table 1. As is shown in Table 1, the "pass" percentage is 82%.

Classification	#Photos
Perfect	4
Good	2
Acceptable	2
Average	1
Bad	1
Awful	0
Failed	1
Pass	9
Fail	2
Total	11
Pass Percentage	82%
Fail Percentage	18%

Table 1: Evaluation of the resulting images created using the Inexpensive Camera support Algorithm (only GPS data, no camera direction available)

8. SUMMARY – FUTURE RESEARCH

We have described SPIM, a system that automatically augments the information in nature pictures with semantic information related to the picture objects and information from external sources. To do that, it employs semantic maps that extend the idea of traditional maps with semantic tags describing semantic concepts and individuals. The system allows contextual information to be visualized on the maps (reflecting the interests of the user) as well as on the pictures. The automatic processing of the pictures requires image registration with the contextual geospatial data. The identification and visualization of the semantic objects of interest to the user is done on a second step using the predicted geospatial coordinates of every point of the image, after the correct image registration. These objects may be too small and difficult to automatically discover in the picture. A number of new algorithms have been used for experimentation purposes.

Numerous extensions are foreseen that take into account additional contextual sensors for improving the performance of algorithms. For the purpose of this application environment we have found useful to have by default included in the semantic maps information on semantic individuals of nature concepts (like mountain tops, sea, rivers, etc.) since they are important for the correct registration of the images taken. In addition, for medium distances of the semantic objects of interest, it may be useful to include information related to the semantic individual's footprint, approximate height and color.

The ability to represent persons (actors) within the images has not been exploited by the system. Portrait pictures and pictures with humans can be identified with current technologies and techniques such as face recognition. They could be used for the automatic or semiautomatic detection of specific persons and their silhouettes or faces.

9. REFERENCES

[1] Bannour, H. and Hudelot, C. 2011. Towards ontologies for image interpretation and annotation. In *Proceedings of the 9th International Workshop on Content-Based Multimedia Indexing* (Madrid, Spain, 13-15 June 2011). CBMI 2011. IEEE, 211-216.

[2] Christodoulakis, S., Kontogiannis, P., Petridis, P., Moumoutzis, N., Anastasiadis M. and Margazas T. 1997. A Distributed Multimedia Tourism Information System. *In Proceedings of the International Conference in Information and Communication Technologies in Tourism* (Edinburgh, Scotland). ENTER 97. 295-306.

[3] Beveridge, J.R. and Riseman, E.M. 1997. How easy is Matching 2D Line Models Using Local Search?, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19, 6 (June 1997), 564-579.

[4] Exif 2002. Exif Version 2.2 Digital Still Camera Image File Format Standard. *Japan Electronics and Information Technology Industries Assoc*. http://www.exif.org/Exif2-2.pdf.

[5] Google Earth. 2013. Google Earth virtual globe. http://earth.google.com.

[6] Salembier, P. 2001. MPEG-7 Multimedia Description Schemes. *IEEE TA on Circuits and Systems for Video Technology* 11(6), 748–759.

[7] Schreiber, A.Th., Dubbeldam, B., Wielemaker, J., Wielinga, B. Ontology-Based Photo Annotation. 2001. *IEEE Intelligent Systems*. 16(3) 66-74.

[8] Smith, J.R. and Chang, S-F. 1996. Visualseek: a fully automated content-based image query system. *In Proceedings of ACM Multimedia 1996* (Boston, US). ACM Press 87–98.

[9] Tsinaraki, C., Polydoros, P., and Christodoulakis, S. 2007. Interoperability Support between MPEG-7/21 and OWL in DS-MIRF. *IEEE Transactions on Knowledge and Data Engineering*. 19(2) 219-232.