DISTRIBUTION OF SEMANTIC ANNOTATIONS TOWARDS A SET OF SPATIALLY-ORIENTED PHOTOGRAPHS

A. Manuel[1,2], C. Stefani[1], L. De Luca[1] and P. Veron[2]

[1]UMR 3495 (CNRS/MCC) MAP, Ecole Nationale Supérieure d'Architecture, 184 Av. de Luminy, 13009, Marseille, France
(adeline.manuel,chiara.stefani,livio.deluca@map.archi.fr)
[2]UMR CNRS 7296 LSIS, Ecole Nationale Supérieure d'Arts et Métiers, 2 cours des Arts et Métiers, 13617, Aix-en-Provence, France
(philippe.veron@ensam.eu)

KEYWORDS: Cultural heritage, photogrammetry, image-based modelling, image processing, semantic annotation

ABSTRACT:

Today with the development of new information technologies, the domain of conservation and promotion of cultural heritage benefits from new tools to disseminate knowledge on heritage buildings. These tools are intended to help experts to enrich and access information about buildings. This paper describes this critical issue and proposes an approach permitting the automatic transfer of annotations in a set of oriented photographs (whose positions and orientations in space are known) by using 3D information. First of all, photographs are processed with an automated image-based 3D reconstruction method to produce 3D information (specifically 3D coordinates). Then, 3D information is used in the process of annotations' transfer between images. Finally, with this process, the modification of annotations can be easily implemented. As a consequence, this process provides a simple way to annotate blocs of images all at once instead of one by one.

1. INTRODUCTION

In the domain of conservation and promotion of cultural heritage, the development of information technologies offers new possibilities to experts for sharing knowledge. The new developed tools permit today to manage a large amount of data and to collect, to structure and to distribute information.

The digitisation and 3D reconstruction methods have became a privileged support for buildings' documentation. The acquisition and 3D reconstruction methods have known significant progress in recent years and permit to create a quite precise representation of buildings, as well as to follow their evolution in time. However, these methods do not always satisfy all the specialists' needs.

In order to study a building, experts use a lot of iconographic sources (drawing, painting, photographs, etc). These sources are very numerous and testify the state of buildings at a specific time. In particular, with the development of digital cameras, the production of photographs is easier. They are supports containing a high level of details in terms of shapes and colors. In this sense, photographs represent an important support of documentation. Furthermore, photographs permit to perform several analyses. Firstly, they can be annotated, entirely or partially, with the help of keywords or ontologies. Secondly, they can constitute a support for carrying out analyses of architectural shapes (measuring, outline extraction, shape recognition …). Thirdly, they can also constitute a support for characterizing buildings' surfaces and for observing the state of conservation (for example, degradation phenomena). Finally, thanks to the progress in photogrammetry, 3D representations can be rapidly generated from images and are as precise as 3D models obtained from laser acquisition.

If, today, photography reveals as an essential way for annotating and analysing the morphology and the state of conservation cultural edifices, a main problem emerges: for an exhaustive documentation of cultural heritage, it is necessary to manage the collect of hundreds or thousands photographs. For the large number of manipulated photographs, the propagation and the distribution of annotations (areas, surfaces, measures ...) among all images should be automatic.

For this reason, studying the transfer of annotations in a set of oriented photographs (which position and orientation in space are known) has been envisaged. The objective of this research is to establish a solution for the transfer of an annotation, associated to a photograph, to the other oriented photographs.

This article has been divided into six sections. Section 2 examines some methods for the annotation of images and 3D models. Section 3 presents the general approach. Section 4 and 5 present respectively the 3D reconstruction method used and the method adopted for the propagation of annotations. Finally, the last section evaluates the system, assessing its limits and fixing some research perspectives.

2. RELATED WORK

In the cultural heritage domain, annotating iconographic sources, and more specifically photographs (2D annotation), or annotating 3D models (3D annotation) gives supplementary information helping the comprehension of buildings.

Nowadays, current researches use three ways for the 2D annotation of images: the manual annotation, the automatic annotation and the semi-automatic one. Manual annotations are defined by the user on images one by one by using either keywords (Halaschek-Wiener C. et al, 2005), or ontologies (Petridis K. et al, 2006). Others researches use automatic annotations (Shotton, J.D.J. et al, 2009 - Akcay H.G. et al, 2008). These methods are based on the analysis of the image content by means of two steps: firstly, segmentation of images and, then, shape recognition. Finally, others methods try to combine manual and automatic methods: the semi-automatic

methods. The combination of the two methods can be done in two ways: by using firstly the manual method and, then, the automatic method (Barrat S. et al, 2009), or by using the automatic method at first and, then, a manual validation of the user like on the website alipr.com (http://alipr.com/).

Regarding 3D annotations, some methods have been developed. Several researches choose to use already existing standards like X3D, MPEG-7 (Bilasco I.M. et al, 2005) or semantic web (Pittarello F. et al, 2006). The most interesting method is the one using a segmentation of the 3D model allowing attaching an instance on each part of the segmentation (Attene M. et al, 2007 - Attene M. et al, 2009).

Finally, other researches combine 2D and 3D information. The annotation can be supported by the picture (Snavely N. et al, 2006) or by the 3D model (Busayarat C., 2010) and uses 3D information that is transferred between images.

Clearly, these works have shown that the process of annotation can be significantly improved, on one hand, by connecting the iconographic collection to the building, and on the other hand, by semantically annotating buildings in terms of their parts and subparts.

However, today the semantic relation between the 3D-model and the collection of spatialized pictures is just at the beginning of its possibilities. The use of semantics annotation could become a support for displaying measurements made on the accurate 3D-model, information collected about analytical data, or still the conservation state of the building.

3. MAIN APPROACH

The main objective of this research is to develop a process for linking 2D annotations in a set of oriented photographs. The approach is based on the idea that 3D information could serve as a support for transferring a 2D annotation, defined on one image, to the other relevant images of the set. Thus, having information on the spatial position of images and on the depth of pixels, it is possible to obtain a relation between images.

For these reasons, the adopted approach permits to:
- generate 3D information by using an automated image-based 3D reconstruction
- transfer annotations from an image to another one, by comparison of pixels' depth.

These two aspects would be detailed in the two next sections.

4. AUTOMATED IMAGE-BASED 3D RECONSTRUCTION

The MAP laboratory, in collaboration with IGN, contributes to the development of a chain of Automatic Picture Processing for Reconstruction of 3D objects in the project TAPEnADe (http://www.tapenade.gamsau.archi.fr/TAPEnADe/). This chain, detailed in (Pierrot-Deseilligny M. et al, 2011a) consists of three axes:
- Image Triangulation
- Surface measurement with automated multi-image matching
- Point cloud generation

The aim of this chain is to automatically calibrate and orientate a set of photographs and to generate very dense point clouds (up to 1 3D point for 1 pixel) of the studied objects.

4.1 Image triangulation

This method is based on the open source APERO (Pierrot-Deseilligny M. et al, 2011b). It is constituted of different modules for tie point extraction, initial solution computation, and bundle adjustments for relative and absolute orientation. This open source is used for the calculation of camera positions. For each image, an XML file (containing the spatial position and orientation of the image) is generated and permits to position images one relative to another.

4.2 Surface measurement with automated multi-image matching

As photographs' positions are known, a dense point cloud is calculated for each master image with the open source MicMac (Pierrot-Deseilligny M. et al, 2006). The matching has a multi-scale, multi-resolution, pyramidal approach (Figure 1) and derives a dense point cloud using an energy minimization function. For each hypothetic 3D point, a patch in the master image is identified and projected in all the adjacent images and a global similarity is calculated.
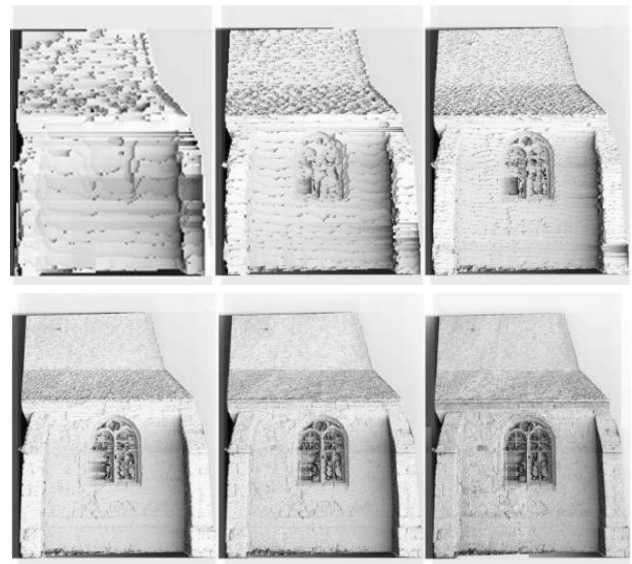

Figure 1: Pyramidal approach: example of results during the multi-scale matching

4.3 Point cloud generation

Starting from the results on the positions and orientations of photographs, as well as the results of the multi-stereo correlation, the depth maps are converted into 3D metric points. This conversion is based on a projection in space of each pixel of the master image taking into account the orientation and position of the image. Thus, a dense point cloud is created for each master image (Figure 2).

Figure 2: The multi-stereo image matching method: the master image (left), the matching result in the last pyramidal step (center) and the generated colorized point cloud (right)

As orientation and position of each image are considered at the moment of the generation of point clouds, the superposition of each point cloud associated to each master image permits to create directly the dense point cloud of the building (Figure 3).

Figure 3: Superposition of point cloud of all master images

4.4 XYZ Files

Using the point cloud generated for each master image, the coordinates in space of each pixel of the image are known. These coordinates can be then stored in a TIFF file.

A TIFF file is an image file storing color information of an image on three layers associated each to a primary color : Red, Green and Blue (RGB). The color of an image's pixel is so a combination of red, green and blue, each of them represented by a value between 0 and 255. Each layer can be considered as an array containing one of the three values of the pixels' color. The association of these three value permits to see the color of pixels (Figure 4).
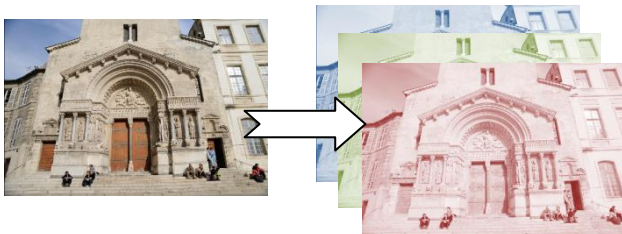

Figure 4: Structure of a TIFF file with the three color's layers: Red, Green, Blue

Instead of storing color information on the image, the created TIFF file will store the 3D coordinates (X, Y and Z) of pixels. X coordinates will be stored in place of the value of Red, Y

coordinates will be stored in place of the value of Green and Z coordinates will be stored in place of the value of Blue. Thus, three arrays are available; each of them containing one of the three coordinates (Figure 5).
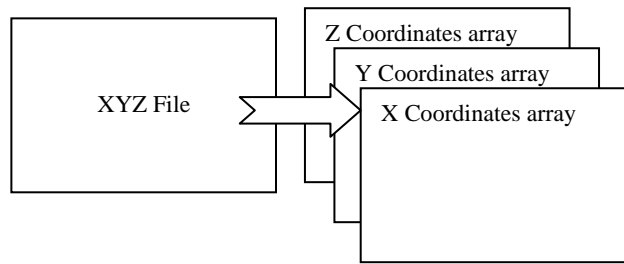

Figure 5: Structure of XYZ files with the three coordinates: X, Y and Z.

Knowing the pixel position in the image (row and column), one of its coordinates in space could be extracted from the associated XYZ file, by reading the value in the array containing this coordinate. This value is at the same row and column as in the image. Thus, in order to find the X coordinate's value of the pixel at the row i and the column j in the image, it is sufficient to find the value at the row i and the column j in the X array of the XYZ file of the image.

5. METHODOLOGY FOR THE PROPAGATION OF ANNOTATIONS

The adopted methodology for the propagation of annotations between images uses TIFF files containing the coordinates.
The propagation consists in three steps (Figure 6):

- Definition of the annotation on one image
- Research of the coordinates X, Y and Z of the annotated area
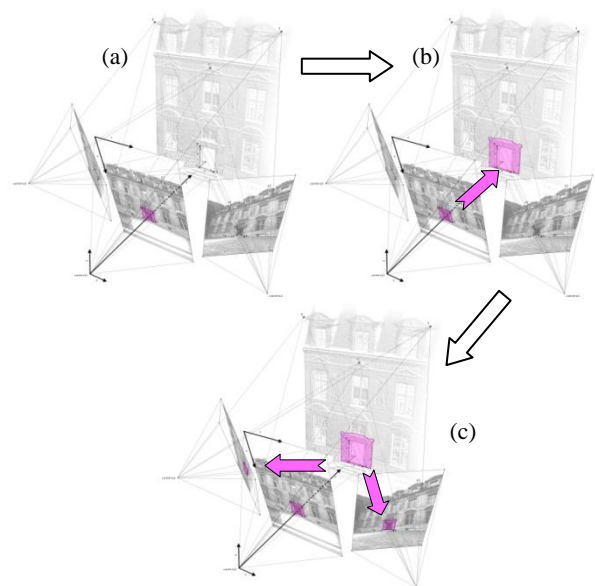- Projection on the other images of the set

Figure 6: Steps of transfer: (a) definition of the annotation on the middle image, (b) research of X, Y and Z coordinate of the area, (c) projection on the other images

Each annotation will not be defined by an area on an image but by a set of triplets of coordinates in space.

5.1 Definition of annotations

The first step consists in defining the area of the annotation. For this, a region of interest (ROI) is drawn on one image of the set. Starting from this drawn area, a mask is constructed: it is as wide as the image and contains white areas on a black background. The white area corresponds to the position of the drawn ROI (Figure 7).
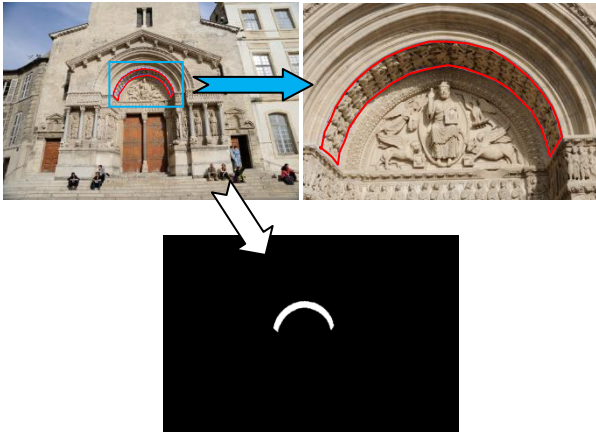

Figure 7: Definition of the ROI (up) and extracted mask (down)

5.2 Research of the 3D coordinates of the area

In order to extract the X, Y and Z coordinates of the annotated area, the position (row and column) of pixels of this area must be known. In this way, for each white pixel of the previous mask, the position of the pixel is searched and a list of couples of values i (row) and j (column) is constructed. Having the position of white pixels of the mask, the reading of X, Y and Z coordinates at the same positions in the XYZ file of the image gives a list of X, Y and Z triplets (Figure 8).
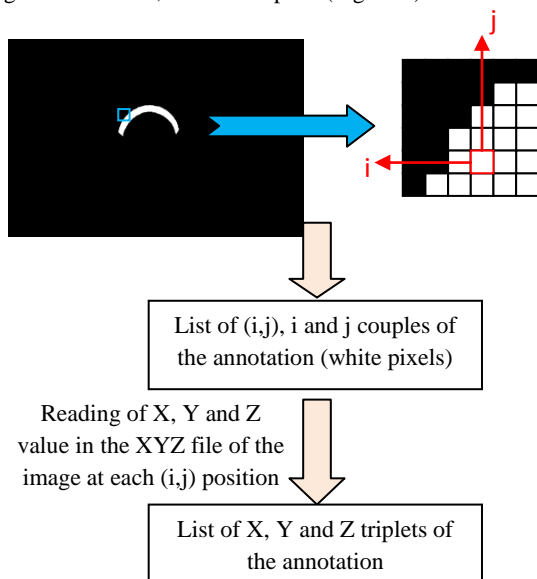

Figure 8: Research of X, Y and Z coordinates of the annotation

This list of X, Y and Z triplets permits to define the annotated area in space (Figure 9). This list will be the data set for the storing of annotation.
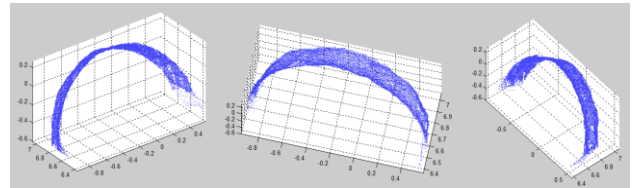

Figure 9: 3D representation of the coordinates list of the annotation

5.3 Projection on the others pictures

Knowing the X, Y and Z coordinates of the points representing the annotation, these points must be retrieved in the other images of the set. For this, the use of XYZ files of each image is needed.

The principle consist in comparing each coordinates' triplets of the annotation with all X, Y and Z coordinates of the XYZ files attached to the image on which the annotation must be transferred (Figure 10).
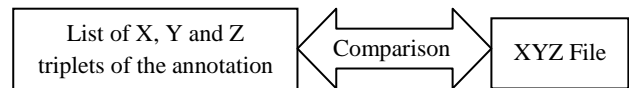

Figure 10: A comparison between the list of X, Y and Z triplets of the annotation and the XYZ files helps to project the annotation on the other images

The XYZ file is scanned in order to test all combinations of positions i (row) and j (column) (that means for i from 1 to the number of rows of the image and for j from 1 to the number of column of the image). For each (i,j) combination, the X, Y and Z values are read in XYZ file and, then, are compared with the annotation's triplets. If one of the annotation's triplets is equal to (i,j) position's triplets (if X values are equal and if Y values are equal and Z values are equal), this position is defined as true. Otherwise, the position is defined as false (Figure 11). Thus, if a X, Y and Z triplets is not finding in the XYZ files, that means it does not appear in image.
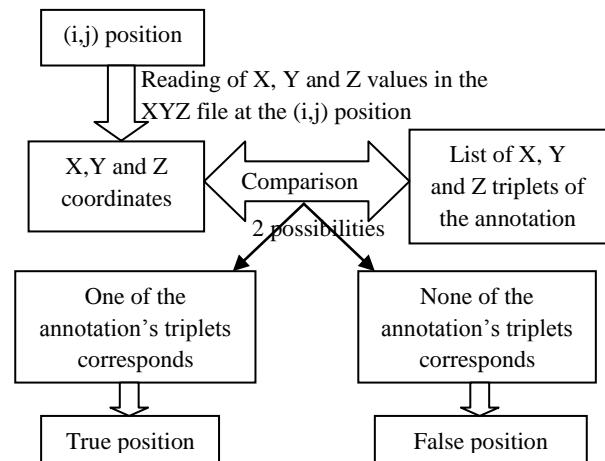

Figure 11: Research of true position and false position on the image

A mask is then constructed by taking into account true and false positions. In the case of a false position, a black pixel is affected in the mask at this position. In the case of a true position, a white pixel is affected in the mask at this position (Figure 12).

```
┌─────────────────────┐        ┌─────────────────────┐
│  (i,j) true position │        │ (i,j) false position│
└─────────────────────┘        └─────────────────────┘
          │                               │
          ▼                               ▼
┌─────────────────────┐        ┌─────────────────────┐
│ (i,j) position on    │        │ (i,j) position on    │
│ the mask : white     │        │ the mask : black     │
│ pixel                │        │ pixel                │
└─────────────────────┘        └─────────────────────┘
```
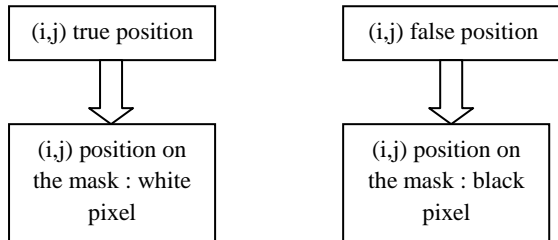
Figure 12: Interpretation of true and false positions for mask's construction

The mask holds a white area corresponding to the detected area. As the XYZ file has the same dimension as the image and as all positions are tested, the mask has also the same dimension as the attached image (Figure 13).
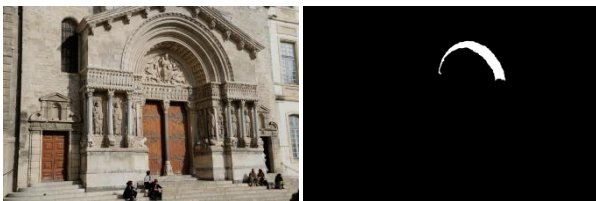
Figure 13: Image (left) and associated mask by the propagation of annotation (right)

If the drawn area (corresponding to the annotation) does not appears in one of the other images, the created mask for this image will have only black pixels.

By affecting a transparency value on the mask and a color value on white pixels, the superposition of the image and of the mask permits to display the annotated area (Figure 14).

Figure 14: Visualisation of the area on another image

In this process, the transfer can be performed in two ways for an image: from the picture (definition of an annotation) or to the picture (transfer from another picture).

5.4 Multi-view enrichment of annotations

As the transfer only allows the search of existing points on the annotated image, in some cases it is necessary to complete the annotation from another view. Indeed, in some cases, the object to be annotated does not appear wholly in any image and needs other views to be completely represented.

For this reason, the objective of multi-view enrichment is to permit the user to define another ROI on another image for completing the annotation and to implement the propagation with the union of the two drawn areas.

A first definition of the annotation is made on a view. By implementing the steps 5.1 and 5.2 on this view, a first list of X, Y and Z coordinates is extracted. Then by implementing again the steps 5.1 and 5.2 on another view, a second list of X, Y and Z coordinates is also extracted. These two lists are associated in order to have only one list of X, Y and Z coordinates. Finally the step 5.3 is processed for all images by using the combination of the two extracted lists (Figure 15).
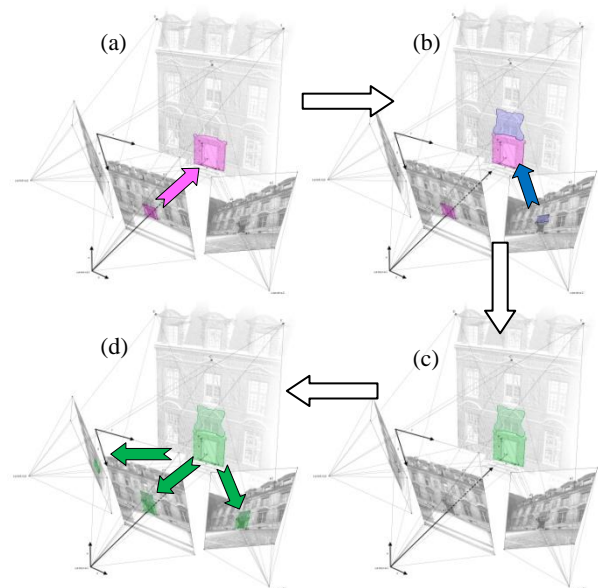
Figure 15: Steps of multi-view definition: (a) definition of a ROI on a view and search of X, Y and Z coordinates (pink), (b) definition of a ROI on another view and search of X, Y and Z coordinates (blue), (c) association of the two list of coordinates (green), (d) projection of the modified annotation on others images

These steps can be generalised for the use of more than two images. Indeed, whatever the number of necessary images for the complete definition of an object to annotate, it is sufficient to draw the wanted area on each images, to search the X, Y and Z coordinates of each drawn area and, then, to assemble all the lists of coordinates before implementing step 5.3.

This enrichment of annotations can be performed from different viewpoints and the definition of an annotation can be defined at best.

6. CONCLUSION AND PERSPECTIVES

This work has described a process based on 3D information to transfer annotation to a set of spatially oriented pictures of a building. Despite the results obtained with this study, some issues need to be resolved and some reflections should prompt further research.

First of all, the results of the transfer of the annotation can be improved. The results show a selection very close to the expected selection. However, near the edges of the picture,

results obtained with MicMac are not very precise and this could cause errors in the transfer of annotations.

Besides, the set of used photographs must be updated. The system should provide the adding of new photographs to the already annotated images.

Then, in order to improve the definition and the transfer of an annotation, a segmentation of the image or of the implicit point cloud (implicit because contained in the XYZ files) could be envisaged.

Afterward, for now, only the definition of the annotated area is possible. The simple annotated area is not useful. A semantic description must be associated in order to improve the understanding of the building.

Furthermore, with the help of images or point cloud, a set of 2D or 3D analyse tools (color, shape …) could be developed.

At last, if annotations are semantically defined, in a future development all data can be crossed according to different criteria permitting to formulate several queries (by single annotation, by terms …).

7. REFERENCES

Akcay H.G., Aksoy S., 2008, Automatic Detection of Geospatial Objects Using Multiple Hierarchical Segmentations, *IEEE Transactions on Geoscience and Remote Sensing*, 46(7), pp. 2097-2111.

Attene M., Robbiano F., Spagnuolo M.et Falcidieno B., 2007, Semantic Annotation of 3D Surface Meshes based on Feature Characterization.

Attene M., Robbiano F., Spagnuolo M., Falcidieno B., 2009, Part-based Annotation of Virtual 3D Shapes.

Barrat S., Tabbone S., 2009, Classification et extension automatique d'annotations d'images en utilisant un réseau Bayesien.

Bilasco I.M., Gensel J., Villanova-Oliver M., Martin H., 2005, 3DSEAM: a model for annotating 3D scenes using MPEG-7.

Busayarat C., 2010, La maquette numérique comme support pour la recherche visuelle d'informations patrimoniales.

Halaschek-Wiener C., Jennifer G., Andrew S., Michael G., Bijan P., Jim H., 2005, PhotoStuff -- An Image Annotation Tool for the Semantic Web.

Petridis K, Anastasopoulos D, Saathoff C, Timmermann N, Kompatsiaris I, Staab S, 2006, M-OntoMat-Annotizer: Image annotation linking ontologies and multimedia low-level features. *KES 2006 10th Intnl. conf. on knowledge based, intelligent information and engineering systems*.

Pierrot-Deseilligny M., Paparoditis N., 2006, A multiresolution and optimization-based image matching approach: An application to surface reconstruction from SPOT5-HRS stereo imagery.

Pierrot-Deseilligny M., De Luca L., Remondino F., 2011a, Automated image-based procedures for accurate artifacts 3D modeling and orthopicture generation.

Pierrot-Deseilligny M., Clery I., 2011b, APERO, an open source bundle adjustment software for automatic calibration and orientation of set of images, *Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38(5/W16), on CD-ROM. *ISPRS Int. Workshop 3D-ARCH 2011*, Trento, Italy.

Pittarello F., De Faveri A., 2006, Semantic Description of 3D Environments: a Proposal Based on Web Standards.

Shotton, J.D.J., Winn, J., Rother, C., Criminisi, A., 2009, TextonBoost for Image Understanding: Multi-Class Object Recognition and Segmentation by Jointly Modeling Texture, Layout, and Context, *IJCV*(81), No. 1.

Snavely N., Seitz S., Szeliski R., 2006, Photo Tourism: Exploring Photo Collections in 3D.