ANALYSIS, THEMATIC MAPS AND DATA MINING FROM POINT CLOUD TO ONTOLOGY FOR SOFTWARE DEVELOPMENT

R. Nespeca a, *, L. De Luca b

a DICEGA, Dept. of Civil, Building Engineering and Architecture, Engineering Faculty, Polytechnic University of Marche, Ancona, Italy – r.nespeca@univpm.it

b MAP (UMR 3495 CNRS/MCC) - Modèles et simulations pour l’Architecture et le Patrimoine, CNRS, Marseille, France - livio.deluca@map.cnrs.fr

Commission V, WG V/2

KEY WORDS: Analysis, Thematic Maps, Point Cloud, Cultural Heritage, Degradation Recognition, Restoration

ABSTRACT:

The primary purpose of the survey for the restoration of Cultural Heritage is the interpretation of the state of building preservation. For this, the advantages of the remote sensing systems that generate dense point cloud (range-based or image-based) are not limited only to the acquired data. The paper shows that it is possible to extrapolate very useful information in diagnostics using spatial annotation, with the use of algorithms already implemented in open-source software. Generally, the drawing of degradation maps is the result of manual work, so dependent on the subjectivity of the operator. This paper describes a method of extraction and visualization of information, obtained by mathematical procedures, quantitative, repeatable and verifiable. The case study is a part of the east facade of the Eglise collégiale Saint-Maurice also called Notre Dame des Grâces, in Caromb, in southern France. The work was conducted on the matrix of information contained in the point cloud ascii format. The first result is the extrapolation of new geometric descriptors. First, we create the digital maps with the calculated quantities. Subsequently, we have moved to semi-quantitative analyses that transform new data into useful information. We have written the algorithms for accurate selection, for the segmentation of point cloud, for automatic calculation of the real surface and the volume. Furthermore, we have created the graph of spatial distribution of the descriptors. This work shows that if we work during the data processing we can transform the point cloud into an enriched database: the use, the management and the data mining is easy, fast and effective for everyone involved in the restoration process.

1. INTRODUCTION

The survey for restoration is characterized by an increased focus on reading the state of conservation of the building, both architectural and archaeological. The traditional architectural drawings are always accompanied by graphical thematic representations, like the survey of constructive and decorative details, the masonry layering, the crack pattern, the degradation and the color recognition. The drawing of these maps is usually manual work by skilled operators who often rely in situ analysis, on the photographic acquisitions and on their knowledge and skills. It is very important the phase of the investigation in situ, where most of the visual or instrumental investigation work is concentrated. Often, however, the record media and investigative techniques are very traditional, slow and difficult to reproduce: notes that are to be digitized and direct observation as the only means to recognize degraded areas. Therefore, the qualitative and subjective investigations anticipate the step of geometrical and instrumental survey, like if they are two separate and independent stages.

The scientific issues concerning the digital data structuring and sharing have great applications in the field of conservation. Indeed, the management of the conservation and restoration activities involves several actors such as administrators, practitioners and scientists (curators, microbiologists, chemists, geologists, materials scientists, architects, historians and engineers). The diversity of experts involved in a restoration project therefore requires, first, bringing together and archiving in a consistent and structured way the heterogeneous data produced by each actor. Methods and common tools for collaborative decision support, which are tools for the dynamic processing crossing various disciplines, must be designed and implemented in order to explore new digitally driven ways for enriching the decision-making process in conservation. (Clini, et al., 2015)

The diagnostic objective is to assess the current state of the building and makes use of destructive mechanical tests, performed on samples taken from the structure as undisturbed as possible and non-destructive testing (non-invasive), which do not cause any damage to the artefact. (Wittich, et al., 2012)

The extractable information from the geometric and topographical survey using the new technologies are an essential support to the next step of analysis, data integration, diagnostics and rehabilitation project (Monni, et al., 2015). The point cloud is itself a morphometric data storage that if properly treated can provide guidelines to diagnostic investigations.

The creation of three-dimensional representations of heritage artefacts requires methodologies to digitally acquire, to interpret and to describe the morphological features and the visual appearance. (De Luca, et al., 2011)

The work presented here is part of the field of research conducted by the MAP laboratory, mixed research unit of the CNRS at Marseille, about the 2D-3D semantic annotations and creating ontologies (De Luca, 2014). In particular, this work fits perfectly into the MONUMENTUM, an on-going research project funded by ANR (French National Agency for Research) which integrates works of CNRS-MAP and IGN-MATIS concerning the image-based-modelling of architectural heritage and works of CNRS-MAP, CRM and CICRP focusing on the development of a 3D...
The study is focusing on development of software platforms, preferably open source, allowing the management and sharing of digital data. In this context, some studies (Manuel, et al., 2013) allow complex interactions, semantic annotations, statistical surveys and distributions of phenomena for the diagnosis. Other studies (Fassi, et al., 2015) developed solutions that favour the sharing of information in real-time, such as to allow changes or annotations on the field and instantly shared to other team members. Both solutions provide for the use of the three-dimensional model.

Increasing information on a building goes through the annotation of iconographic sources and more specifically photographs (2D annotations) and 3D representations (3D annotations). Annotations on 2D sources can be set up by means of three main methods: manual, automatic and semiautomatic annotation methods. With manual methods, images are annotated one by one in their entirety or partiality.

Keywords or ontologies can be associated to the annotated parts. Automatic methods automatically assign description to images by means of an analysis of image content. Finally, semiautomatic methods use automatic methods’ processes but ask also for manual intervention.Annotations on 3D representations consist in attaching annotations to parts of the 3D representations i.e. to points, to segments, to surfaces or objects in the scene. In past years, the use of 3D information for images annotation has known an increased interest. All works show that there is an important interest in connecting images with 3D representations for the process of images annotation. Other work presents the first principles for the development of an information system to monitor the historic building degradation based on three main components: a HDR image-based 3D digitization automatic pipeline, a hybrid (2D/3D) semantic annotation method and a domain ontology describing knowledge related to degradation phenomena. (Messouidi, et al., 2014)

The resulting semantically structured model can also be enriched by the definition of the model behaviour of components, material parameters, boundary conditions and the initial state (Bagnéris, et al., 2013)

In the following paragraphs, it shows how from the point cloud can be released the following quantities:
- the normal vector for each point;
- the luminance value calculated in ambient occlusion;
- the curvature of the mesh surface calculated ad hoc;
- the roughness in relation to an area of interest.

After the point cloud enrichment by these geometric information we proceeded to the study of selection algorithms that operate directly on these surface descriptors until you get to more advanced clustering methods. It’s very interesting the possibility of selecting areas, with automated processes, that have the same value of roughness, calculated according to mathematical quantitative, repeatable and controllable procedures. If the roughness value corresponds to a certain surface corrosion state, it’s possible to extrapolate thematic maps with semi-automatic and quantitative procedures.

2. CREATION OF 3D MODEL OF THE CHURCH OF CAROMB

The case study is the Eglise collégiale Saint-Maurice also called Notre Dame des Grâces at Caromb, in the south of France. The church was the subject study of the master thesis in “Rehabilitation et Sauvegarde du Patrimoine Architectural”. The previous studies were offering exceptional database and analysis procedures for the validation of experience in this work. Cartography available had been produced by the visual and photographic surveys.

The maps were:
- the classification of the stones according to the colour and granulometry; (Figure 1)
- the recognition and classification of the carving;
- the identification of the constructive abnormalities depending on size and distribution of the blocks;
- the recognition of previous restorations due to the presence of new materials or signs of subsequent shots;
- the identification of areas affected by biological colonization according to the colour change; (Figure 2)
- the creation of maps of detachment, corrosion or loss of material based on the recognition of areas of depression or with particular characteristics of surface roughness. (Figure 3)

Figure 1. The classification of the stones according to the color and granulometry.

Figure 2. Maps of the areas affected by biological colonization.

Considering the traditional investigations carried out and previously shown, the idea at the heart of this work is to research extraction semiautomatic procedures of similar data and to facilitate their management. The investigated features are about geometry and materials. These analysis requirements justify the methodology and operations described in the following paragraphs. This cartography are based on the chromatic variations and alterations on the surface of the stones of the wall. A photographic and geometric acquisition, whatever the technique used, give a point cloud full of raw information. The RGB values and a good density of the cloud can provide the necessary information.

The previous thematic maps were made on photographic images used for the reconstruction of three-dimensional model through the photogrammetric technique. To narrow the area and decrease the computational effort we chose an extract of the eastern façade already investigated and analysed with traditional cartography. The starting point is the creation of the point cloud by MicMac
open-source software. From a previous acquisition campaign, 18 photos have been selected. (Figure 4)

Figure 3. Maps of the detachments, corrosion and loss of matter.

For scaling, geo-referencing and validating the photogrammetric model, 12 point were take on the point cloud acquired in another campaign by a laser scanner instrument, which they will become the GCP in the photogrammetric model. With open-source software CloudCompare, we have extracted the coordinates chosen in a format .txt, then converted into MicMac in .xml. (Figure 5)

Figure 4. Orientation phase.

The next step is the recognition of the extracted point on photographic images and automatic adjunct of the coordinates, allowing the scale and rotary translation of the photogrammetric model coherently with the laser scanner system.

Figure 5. Extraction of the coordinates of points, and conversion of the list for the scale, geo-referencing and validation of the photogrammetric model.

The starting point was:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>z</th>
<th>r</th>
<th>g</th>
<th>b</th>
<th>nx</th>
<th>ny</th>
<th>nz</th>
</tr>
</thead>
</table>

From these values, it is possible to work on the existing descriptors, such as normal, and the extraction of new geometric descriptors can provide additional information, again by open-source software. A preliminary step of decimation with MeshLab has made to ease the computational effort. This phase transform the initial point cloud containing more than 7 million points in a point cloud containing 1.2 million, by an algorithm that would make the regular mesh with a pitch between the points of 5 mm. This is an essential step for the following consideration.

3. DIGITAL MAPS AND DATA MINING

Obtained the point cloud, decimated and clean as needed, it is necessary to compare the photogrammetric model with the laser scanner point cloud for the quantitatively assess of deviations. It emerges from the map a good result in the creation of the geometry, maybe a little worse than a slight rotation which produces a maximum deviation of 10 mm at the edge, probably due to the georeferencing. (Figure 7)

This distance value (C2C cloud to cloud) between the two discrete models is in addition to the listing of the point cloud with another column: a new value at each point.

Figure 6. Dense Image Matching phase.

Figure 7. Map of deviation between photogrammetric cloud and scanner cloud.
Subsequently, using the algorithms implemented in CloudCompare, a plan is generated through RANSAC algorithm. The RANSAC algorithm is an iterative algorithm that allows obtaining a good geometry of fitting on the cloud, enabling the choice of certain calculation parameters such as: the maximum distance from the primitive, the sampling resolution, the maximum angular deviation of the normal and a probability value of overlap. Then, we are created a deviation map between the point cloud and RANSAC plane, adding the listing of C2M value (Cloud to Mesh) (Figure 8).

The extractable information from this map are immediately apparent. All surface imperfections are clear and quantitatively measurable. The manipulation of the color range also gives the possibility of displaying very important for analysis, allowing to select the data and to make them visually communicative.

Subsequently, the map in Ambient Occlusion (AO) is realized, experimenting with different settings and resolution values. The calculated value for each point is the luminance. The range is from 0 to 1, which correspond to the black and white respectively. This size offers two main advantages: a single calculation phase and, very important, the independence from the reference system or a reference geometry. (Figure 9)

In the case of a masonry like the one in question, it emerges very clearly the areas of depression, the joints between the segments or the detachment and corrosion areas, with the possibility of graphic representation of the data obtained for subsequent evaluations.

By overlaying the display of the luminance value and the RANSAC plane, differences and correspondences between these quantities are clear. (Figure 10)

The luminance value in ambient occlusion highlights the convexity of the surface with shades of grey, but shows similar values (equal to 1) on the concave parts. Starting from the components of the normal, already calculated during the creation of the point cloud, the normal map were made. Within the CloudCompare software, this process occurs in two steps:
- a first transformation of normal in gradient directions (the vertical component Nz translates in the Dep value with values between 0 and 90, while the Nx and Ny components determine the value of dip Direction with values between 0 and 360);
- the second phase transforms the gradient directions in the HSV color space values.

Then, three normal maps are displayed: one with the values of Dip (shows the changes in the vertical plane), one with the values of Dip Direction (shows the variations in the horizontal plane), and one in which the values are combined. (Figure 11)

Then, these new RGB values can be saved and added to the listing of the point cloud, like other descriptors. If not present, several programs offer the possibility to calculate the normal to the point. Their generation in post-production is not always easy, it is very important to choose good parameters. A test about it was made with MeshLab. What has emerged is that the software (in this case MicMac) or the laser scanner for the generation of point cloud provide normal values much more precise and coherent with the geometry.

Also not wanting to work on the point cloud like is being done in this work, the good normal vector is basic for the creation of the mesh surface, and its control is an standard operation to evaluate the goodness of the point cloud and to facilitate subsequent elaborations. The next step was the calculation of the curvature value. Initially set a kernel value, then set the size of the survey area around each point, the curvature is estimated using the best fitting of a quadric on the area choice. Three curvature values can be calculated: Gaussian, Mean and Normal Change Rate. Different kernel values were tested for each type of computable
curvature, in order to bring out visual information characterizing the wall. (Figure 12) The best result was obtained with a kernel value equal to 2cm, effectively congruent with the minimum features of this artefact and with the decimation grid set at the beginning of the work (5mm). With the same range, the map which shows more information seems to be that calculated by Normal Change Rate method. (Figure 13).

Finally, the last calculated descriptor is the roughness. For each point, the roughness value is the distance between the given point and the best plan fitting, calculated on nearby point. Then, also in this case, the kernel factor imposed at the beginning of the calculation results to be decisive for the calculation. By varying the kernel, the information vary: the higher the value, the greater the portion of comparison and consequently the geometrical considerations concerning a local scale, but large; the small kernel values seem to show more details, useful for smaller-scale analysis. Given the size of the object, the kernel equal to 200 (20cm) shows very interesting results: in addition to highlighting the depressed areas, the lines of stormwater runoff are defined very precisely. (Figure 14)

For a first attempt to qualitative analysis, tests were made on the color change of the display scale. Choosing the scale more representative of each descriptor grey, the maps were overlaid with image editing software, Photoshop, with the overlay method brighten up. The result is a hybrid map where we can not make quantitative assessments, but we can qualitatively read all information from the individual thematic maps (Figure 15):
- the areas with significant loss of matter;
- the spaces between the blocks of stone;
- areas prone to horizontal surface corrosion;
- depressed areas;
- the rainwater drain lines.

The advantages of these analyses are now clear:
- the extensive automation of cartography;
- the values are calculated on the point cloud, so three-dimensional;
- do not need the modeling phase;
- the extracted descriptors are objective values, measurable and repeatable.

Through this data mining, the point cloud becomes a real three-dimensional multi-data information system, which allows switching from raw data to intelligible information. These variables have proven successful in the definition of the conservation status of the artefact, helping the reading of degradation.

The result of this phase is the point cloud enriched with four new descriptors. (Figure 16)

4. METHODS OF SELECTION AND CLUSTERING
This part of work has been developed with the MatLab software. The first step is to import the point cloud. We have decided to transform the listing from .ply to table. The .ply format consists of a haeder with specific information of the variables contained in the columns, followed by the point listing and the corresponding values. The first elaboration was the visualization of the point cloud: the generation is slow and difficult to manage; the interaction almost impossible.
Given the difficulty of displaying, at the end of each process, we decided to convert the results of operations tested to .ply format. To do this, it was necessary: to save the cloud obtained in .txt; to create a header .mat in which change from time to time, automatically, the number of dots; then to convert it in .txt and finally, to merge these two files, saving a new file in .ply format. Following, we tested the opportunity to select the points in a roughness, we can to generate very interesting operation with a great potential.

By acting on the other parameters, such as color, distance from RANSAC plane or roughness, we can to generate very interesting automatic selection of certain factors and ranges is an operation with a great potential.

About a wall such as the one of this case study, made from stone blocks categorized according to their colour differences, the drawing of a material map, automatic and objective, is allowed by selecting all points with a given RGB value, a certain interval or a recognizable pattern due chromatic juxtaposition.

The Figure 17 shows the selection of the points with a certain value: the result shows that the points are belonging to the joints between keystones and shows an extended area in which the stone colour has undergone a major color change.

![Figure 17. Selection of point cloud, starting from a given value of \(r\) (red) and a selectable range.](image1)

Next, in Figure 18, we can see two other selections by values, respectively, by C2M (RANSAC distance from the plane) and R (roughness). With these procedures, it is easily identified a particular degradation phenomenon that causes corrosion, detachment or loss of material.

![Figure 18. Selection of point cloud, starting from a given value of C2M (cloud to mesh) (left) \& R (roughness) (right) and a selectable range.](image2)

### 5. CALCULATION OF SURFACES, VOLUMES AND DISTRIBUTION ANALYSIS

In this part of the work, we will be presented some calculation algorithms of surfaces and volumes useful knowledge, diagnosis and restoration of the artefact.

The bounding box is one of the most common objects in the 3D model management sw. It is the smaller box containing the object. Then, the first written calculation algorithm is the volume of the bounding box, following identification of the automatic minimum and maximum point, respect to x and y axes (Figure 19).

In the case of closed objects, or with a reference system coherent with the main directions of development of the object, this value can be a maximum parameter for a cataloging based on the size.

![Figure 19. Calculation of volume of the bounding box.](image3)

The point cloud is a discrete model, i.e. a three-dimensional non-continuous model consisting of only points. For this reason, it seems necessary to switch at continuous surface for to calculate the real surface of an area of an object. The solution presented below attempts to solve this problem. If we consider the point as a unit surface area, planar and infinitesimal related to the entire object, the idea is to calculate the total real area by multiplying the unit surface area by the number of points.

The axioms that determine the effectiveness of this calculation are:
- the uniform decimation of the point cloud in the three dimensions;
- the density of the point cloud consistent with the minimum feature.

However, not being an exact, but an estimate calculation, some approximations have to be accepted. These approximations prove to be technically adequate and not incisive for the purposes of the goodness of the result:
- the first is about the unit surface area, from quadric to planar;
- the second depends of the decimation algorithms of the point cloud.

Greater the density of the point cloud, smaller the error made in the first approximation. In addition, we need to subtract the half area of the boundary point to refine the calculation. From the test about the different algorithms and software, it shows that the best algorithm uniform decimation, useful to the calculation of the area as described, is within the commercial software, Geomagic. A first test shows that the error is compatible with the decimation, the density and the characteristics of the point cloud. (Figure 20) Combining the methods of selection presented in the previous paragraph and the algorithm of surface calculation, we can automatically extract the point having certain characteristics and their real surface. In Figure 21 shows an example of the C2M parameter selection, and the calculation of the real surface and the volume of mined areas.

In a [0-1] representation of the C2M value, in which the positive
values assume the white color and the negative values is the black, it clearly emerges how the points with positive value constitute the areas of the masonry whit loss of material. In addition, if we assume the RANSAC plane as the reference plane, having selected and isolated points with a positive value, the volume of material loss is estimated by multiplying the unit surfaces of point for their distance from the plane.

Another interesting analysis is to understand the distribution of new descriptors in space. For example, the graph in Figure 22 shows the distribution of the roughness value along the z-axis. The highest values of roughness belong to those areas prone to loss of material and surface corrosion. Understand the distribution in height is of great help to understand the causes of deterioration.

6. CONCLUSIONS

The comparison with the traditional cartography demonstrates the effectiveness and the reliability of this approach: the deviation map from a fitting plane perfectly describes the areas with loss of material (Figure 23); the roughness map shows the levels of surface corrosion highlighting the preferential lines of storm water runoff.

Another validation is the comparison of the point cloud resulting from the selection of a given neighbourhood of R (red) with the thematic map of the deterioration due to biological colonization. In fact, this type of surface alteration causes a very noticeable colour change; in this case, it is well represented by the variation of the red value (Figure 24).

Forward, the data mined can be merged in a software for spatial annotations of point clouds where, after manual or automatic selections, the descriptors calculated, the spatial distributions and the statistical analyses generated can be extracted.
One of the advantages of this methodology is the choice of operating on the point cloud, which allows splitting the phases of analysis and the data mining from the acquisition technique used. In addition, this work shows that if we work during the data processing we can transform the point cloud into an enriched database: with the implementation software, the use, the management and the data mining is easy, fast and effective for everyone involved in the process of restoration, even if not expert of 3D data handling. (Figure 25)

ACKNOWLEDGEMENTS


We also would like to thank Philippe Bromley, responsible for the Caromb's Church Case Study within the ANR MONUMENTUM project.

REFERENCES


Messoudi, T. et al., 2014. Laying the foundations for an information system dedicated to heritage building degradation monitoring based on the 2D/3D semantic annotation of photographs. EUROGRAPHICS Workshops on Graphics and Cultural Heritage, pp. 1-4.
