PCA-based Localization of Pathology on Heritage Building Masonry

Andreas A. Sidiropoulos

Laboratory of Geodesy&Geomatics, Department of Civil Engineering,
Aristotle University of Thessaloniki, Thessaloniki, 54124, Greece,
sidiropoulosansdreas@gmail.com

Konstantinos N. Lakakis

Laboratory of Geodesy&Geomatics, Department of Civil Engineering,
Aristotle University of Thessaloniki, Thessaloniki, 54124, Greece,
lakakis@civil.auth.gr

Vasiliki K.Mouza

Master of Science on Protection, Conservation and Restoration of Cultural Monuments
vmouza@gmail.com

Abstract: *Laser scanning is one of the most common techniques for the geometric documentation of historical buildings. The last decades, laser scanning has replaced or complemented traditional topographicmethods to a significant degree. The advantages of laser scanning that led to its high appreciation of the scientific community is that it is a fast method of geometric documentation that combines the accuracy of topographic and the wholeness of photogrammetric methods. The data that a laser scanner provides to the user is a large set of XYZ coordinates that is called point cloud. The point cloud needs to be processed so the information that it contains can be available and measurable. In this paper, the Principal Components Analysis (PCA) is being implemented to localize and extract pathology existed on planar walls of heritage buildings. PCA is a statistical procedure of identifying patterns in data. Also, it is being used to analyze data through dimensionality reduction and thus extracting information which is not immediately obvious. The building that was selected for the pathology localization is the GaziEvrenos Baths located at the city of Giannitsa in Greece.*

Index Terms: *Heritage, pathology, point cloud, principal components*

1. INTRODUCTION

THE conservation of heritage architecture monuments is a long procedure that requires the cooperation of scientists from various fields to solve management, engineering and restoration problems. The most common scientific disciplines that work on a restoration project are archaeologists, architects, engineers and surveyors. Usually, the data that all scientists use for their research related to the monument are being provided by the surveyors. So, surveyors implement all the available methods and equipment to provide engineers and archaeologists with as more efficient and accurate information of the building as possible. The main categories for geometric documentation of human heritage are the topographic and photogrammetric techniques. Laser scanning combines the accuracy that is being provided by topographic methods and the wholeness of a picture that is available using photogrammetry.

Laser scanners are devices that are able to measure and store distances and angles in a very high rate of about one million measurements per second. There are various methods of how a laser scanner operates but the main principle is that the device emits a pulse towards the object or area of interest, which returns to the instrument after being reflected on its surface. A clocking device measures the time between the emission of the pulse and its reception from the instrument. In this way the distance instrument-object is being calculated and stored. Simultaneously, the horizontal and vertical angles at the emission time are being stored. Next, the 3D XYZ Cartesian coordinates are being calculated for every reflection on the object which is being described by a point. The total of points that describe the whole scene is called point cloud. Additional to the geometric information, most scanners provide RGB information for each point using internal or external photographic camera.

The fact that laser scanners are able to measure that large amounts of data necessitates the development of tools that can extract information desirable by the user. Principal Components Analysis (PCA) is a statistical procedure that highlights information which exists in data. It is a useful tool that is being used to calculate the normal vector of point clouds. The definition of normal vectors for every point of a set, gives the knowledge of curvature changes on surface curves. The fact that the most common form of building masonry is planar, leads to the thought that every change on curvature is a possible location of pathology. So, the points whose normal vector differentiates from the others describe the outline of cracks and lack of material. The point cloud measurements for this paper were made on the GaziEvrenos Baths, a heritage building located at the city of Giannitsa in Greece.

2. RELATED WORK

An important phase during the conservation of historical buildings is the evaluation of pathology which may lead to structural problems.

Cracks and lack of material must be spotted on buildings' masonry, evaluated and restored. Laser scanning and close range photogrammetry techniques were implemented by [1] to record pathology on façades. Close range photogrammetry was used at parts that laser scanning point clouds, despite the high density of points they contained, were not able to provide good visualization for drawing. Active and passive sensors were used by [2] for radiometric detection of pathologies in historical buildings. A laser scanner, a multispectral camera and a digital SLR camera were used to detect and quantify pathology and the results of the methods were compared. The authors conclude that the best solution is a combination of them having laser scanner as the primary choice. Data acquired by two different laser scanning systems were used by [3] to assess pathologies in façades. The extraction of orthoimages led to recognition of pathology such as humidity and fissures. The geometric documentation using a laser scanner and the creation of sections aided [4] to evaluate deformations and damages on the gothic church of Villamoron (Burgos, Spain). The sections were transversal, longitudinal and horizontal across the building.

3. EXPERIMENT

A.Methods

The method that is being described is mainly based on the information that derives from the normal vectors of point clouds. The normal vectors are being calculated by the Principal Components Analysis (PCA) combined to the technique of the k-nearest neighbors.

PCA is a non-parametric statistical method that reducing dimensionality reveals sometimes hidden information. It is a linear transformation of the original data to a new basis [5].

A normal vector of a plane is the geometric entity that specifies the orientation of the plane. If P is a plane in space R^3 and $P_0 = (x_0, y_0, z_0)$ a point on that plane, then the non-zero vector $\boldsymbol{n} = (a, b, c)$ which is vertical to P it is called normal vector (Fig. 1). For a random point (x, y, z) on plane P, the vector $\boldsymbol{r} = (x - x_0, y - y_0, z - z_0)$ is also on plane P and $\boldsymbol{n} \cdot \boldsymbol{r} = 0$ [6].
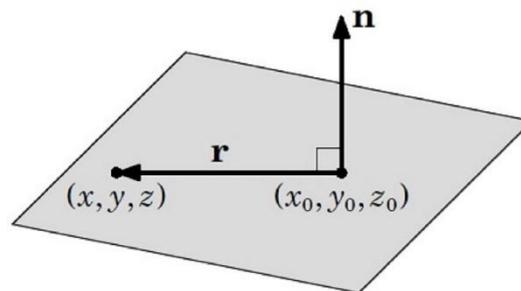

Fig. 1. *Normal vector on plane at point $P_0 = (x_0, y_0, z_0)$ [6]*

Normal vector defined at a point on a surface is the normal of the tangent plane to the surface at that point. The estimation of normal vectors across a surface provides the information of the surface's curvature. The concept of defining normal vectors for points in point clouds is related to the idea of neighborhood and the k-nearest neighbors of a point. This means that calculating the k closest neighbors to a point, a planar surface can be defined and the normal vector of this surface can be

estimated. Under this thought the estimation of the normal vector at each point of the point cloud is feasible. The metric that defines the proximity among points is usually the Euclidean distance. Depending on the application there are various available distance metrics.

B. PCA for normal vectors estimation

PCA offers a method of estimating normal vectors for points in point clouds. Before implementing PCA though, consideration must be given to the value of the number k of the nearest neighbors to a query point q that the normal estimation is required. The definition of this value depends on

1) Scanning density

2) Point distribution

3) Scanning accuracy

4) Curvature of the surface

It is clear that the choice of the number k depends on the data that the users have at their disposal. If the number of the neighbors is n, then the number of the available points to adapt a planar surface and estimate its normal vector is n+1 (including the point for witch the neighbors have been found) (Fig. 2).
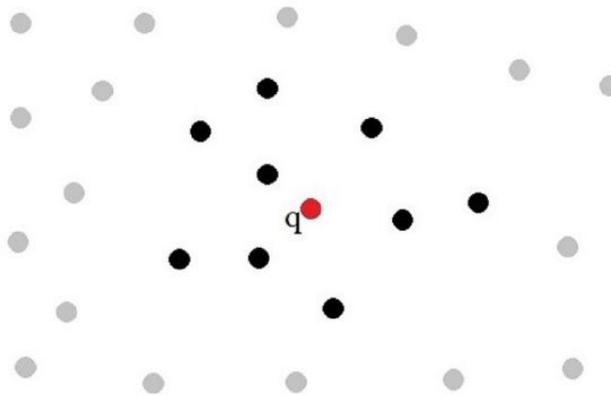

Fig. 2. *k nearest neighbors of point q. In this case, k=9 and the available points for normal vector estimation is 10.*

Once the selection of the k nearest neighbors has been completed for every point in a point cloud, the Principal Component Analysis can be used for the estimation of normal vectors. For generality, we suppose that for every point q of the point cloud, the number k equals to n-1 and the available points for plane fitting is n. So, the data at the users' disposal for the normal vector estimation at point q can be written as

$$X_i(x_i, y_i, z_i) = \begin{bmatrix} x_1 y_1 z_1 \\ x_2 y_2 z_2 \\ \vdots \; \vdots \; \vdots \\ x_n y_n z_n \end{bmatrix}, i = 1,2,\dots,n \tag{1}$$

Each row of X contains the variables and each column the available number of observations for the variables. Using (2) we calculate the mean values for every variable separately.

$$\mu_x = \frac{\sum_i^n x_i}{n}, \mu_y = \frac{\sum_i^n y_i}{n}, \mu_z = \frac{\sum_i^n z_i}{n} \tag{2}$$

The procedure requires the data to be centered around the mean value. So, from every value, the subtraction of the respective mean value is needed. Next, the variance for each variable and the covariance for each pair of variables is being calculated for the mean-centered data by (3) and (4).

$$var\; x = \frac{1}{n-1}\sum_i^n (x_i - \mu_x)^2 \tag{3}$$

$$cov\; xy = \frac{1}{n-1}\sum_i^n (x_i - \mu_x)(y_i - \mu_y) \tag{4}$$

The 3x3 variance-covariance matrix is being filled for the points as shown by (5).

$$C = \frac{XX^T}{n-1} = \begin{bmatrix} var\ x & cov\ yx & cov\ zx \\ cov\ xy & var\ y & cov\ zy \\ cov\ xz & cov\ yz & var\ z \end{bmatrix} \qquad (5)$$

The principal components are the eigenvalues of the matrix C. The three eigenvalues l_1, l_2, and l_3 correspond to three eigenvectors v_1, v_2, and v_3. In the way that the problem has been defined and the data approximate the surface of a plane, the first two eigenvalues will be large and the third one will be small. If this is the fact, the two first eigenvectors lay on the plane that the points are located and the third eigenvector, which corresponds to the smallest eigenvalue, is the normal vector defined by the set of data.

According to the procedure so far, the user can estimate the normal vector for each one of the points in a point set, isolating neighborhoods of k closest neighbors that approach the surface of a plane. In the next paragraph we will discuss the way that the user can exploit the information derived by normal vectors and localize cracks and fissures on flat walls of heritage building masonry.

C. Normal vectors information extraction

Normal vectors indicate the orientation of point clouds at individual points. At the case of planar walls, the calculated normal vectors, with slight divergence, should direct at a mean orientation. That mean orientation in general is the orientation of the whole wall. This fact provides the ability to compare every individual point's normal vector to the mean vector of the plane that depicts the wall. The vectors that diverge more than a specific value from the mean vector indicate possible locations of curvature change and thus possible locations of cracks and fissures existence. The implementation of the comparison can be made to the entire wall, or at specific parts of it that may fit a plane better or present higher interest for the user. In that way the user can focus on more interesting pathology and neglect smaller cracks that may introduce noise to the general localization.

The comparison between the mean normal vector and each individual vector is made by the calculation of the vectors' dot product. The reason for this choice is that the dot product is closely related to the angle that two vectors create.

The value that works as the threshold between the cracks and no-crackspoints is user based. Arecommended definition is one standard deviation of the calculated dot products. Considering that a point cloud has n points, v_ii=1,…,n is the estimated normal vector at each point and v_{mean} is the mean normal vector calculated by all individual normal vectors then, a point is candidate location for cracks and fissures on the wall if (6) is real.

$$abs(v_i \cdot v_{mean}) < coef * std \qquad (6)$$

The value of "coef" is based on the details of information that the user needs to be captured. If the value 1 is attributed to "coef" then the threshold is one standard deviation of the dot products.

D. Tests of the algorithm on GaziEvrenos Baths' walls

The tests of the pathology detection algorithm made for wall parts of the GaziEvrenos Baths. The GaziEvrenos Baths are located at the Greek city of Giannitsa and the 3D data that have been collected are part of the geometric documentation for the restoration proposal of the monument. The laser scanner that was used is the Faro Focus 3D S120. The clouds that have been measured were taken from 29 stations of the instrument, 15 for the exterior and 14 for the interior of the building (Fig. 3).

Fig. 3. *Point cloud of GaziEvrenos Baths at the Greek city of Giannitsa.*

Despite the fact that the joint 3D point cloud from the 29 stations was created, the tests of the algorithm were made to parts of individual point clouds to avoid very high density and malfunction of the software.

In case of the interior, because of the high density of points, the parts of the walls that have been tested are small enough so as the algorithm operates efficiently. For the clouds of the building exterior there was higher convenience of choosing the size of the parts.

Except for the different size of the tested point clouds, the algorithm was operated using different values of the k number of points' neighbors and for the "coef" number that made the choice more or less rigorous against the point selection.

E. Results

The algorithm was operated for various point clouds of the monument's walls. The tests showed that it is very important to choose the appropriate values of the parameters k and "coef". Table I shows the number of points estimated as pathology for part of the southeast façade of the building(Fig. 4) using 3 different values for each one of these parameters. Value 1 attributed to parameter "coef" indicates the outline of pathology, without introducing significant noise, better than values 1.003 and 0.9 for all k groups of neighbors.

Table I. *South East Facade. Total Number Of Points = 358284*

k	coef	Pathology points
11	1	6028
11	1.003	6183
11	0.9	2818
21	1	21548
21	1.003	23361
21	0.9	4977
31	1	12782
31	1.003	13542
31	0.9	3364

Fig. 4. *Orthoimage of southeast façade. White spots highlight the outline of the pathology detected. The chosen combination of the parameters k and coef is 21 and 1 respectively.*

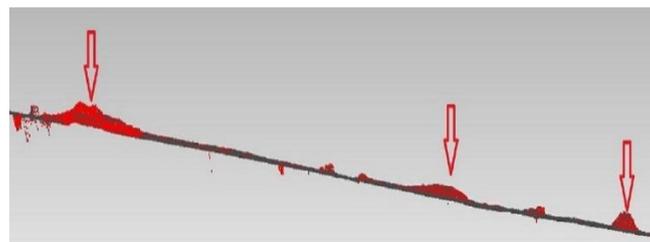If we view this wall part from the top it is easy to locate three significant dents (red arrows at Fig. 5).

Fig. 5. *Top view of southeast façade. Red arrows indicate the locations of cracks on the wall.*

At this locations there is important lack of material. If we go back to Fig. 4, these areas are being highlighted by the three red frames. The points within these frames are those estimated by the algorithm as pathology. Along the façade, in Fig. 4, there are also white points (pathology) that are not distinguished from the top view. These points are also pathology but not that important.

Combinations of k and "coef" that are more rigorous against data do not indicate these areas. So, it is at the convenience of the user and the demands of the project the level of the details that will be exported.

The dimensions of the area in Fig. 4 and Fig. 5 are approximately 6.536m x 2.171m. The part that will be chosen for the pathology estimation may be of smaller or bigger dimensions. It is up to the user and it depends on what are the needs of the application that will be completed.

Fig. 6 shows a test that was completed, this time isolating the large red frame from the the wall shown at Fig. 4. In the case of Fig. 6, the combination of parameters k and "coef" that provide the most satisfying results is 31 and 1.003 respectively. The value and the results of coef =1.003 is very close to the recommended 1. Despite that fact, in other cases of point clouds and pathology the results are very different under these two scenarios.
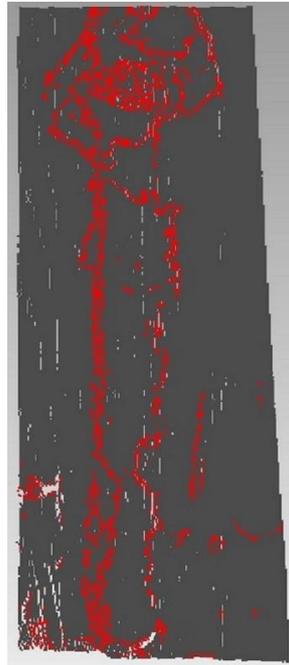

Fig. 6. *Detail from southeast façade. Isolated point cloud, the colored red points (7975 out of 54009) highlight the basic outlines of this wall part pathology.*

Another example from the exterior of the building is shown in Fig. 7. It is a part of the façade that the entrance to the building exists.In that case the main pathology consists of a big hole on the wall's material. The combination of the available values for k and "coef" that gave the best results is 31 and 0.9 respectively. The value 0.9 to the parameter "coef" made able to exclude from the pathology points a big amount of noise.

Fig. 7. *Detail from main façade. Isolated point cloud, the colored red points (8412 out of 42809) highlight the basic outlines of this wall part pathology. Lack of monuments material.*

The most examples that are being introduced refer to the exterior of the GaziEvrenos Baths. This is why the most obvious and characteristic pathology existence is located on the building's façades. The last example of our algorithm implementation refers to the interior of the monument and is shown at Fig. 8. The values of the parameters are 31 to k and 1.2 to "coef".
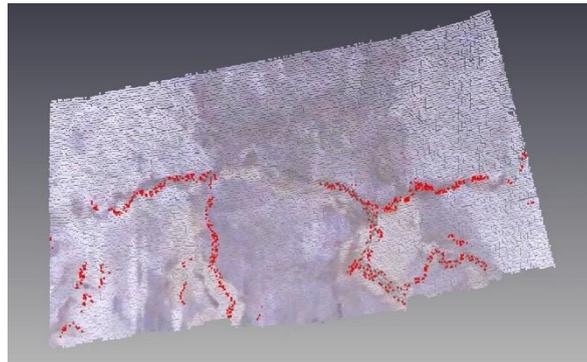

Fig. 8. *Detail from interior of the monument. Isolated point cloud, the colored red points (1532 out of 34322) highlight the outlines of the crack that exist on this wall part.*

4. CONCLUSION

The necessity of localizing pathology on heritage buildings' walls combined to the use of laser scanners for geometric documentation of monuments, arise the idea of creating tools that exploit the information provided by point clouds and export information on the weaknesses that exist on building masonry. This paper introduces such a tool that uses the Principal Components Analysis to estimate normal vectors for the points and with the notion of neighborhood among points in a point cloud aims to export pathology information on flat walls. The results of this tool's implementation show that it is efficient to highlight cracks, fissures and lack of material.

The highlighted points that are categorized as pathology points need further elaboration and continuous examination by the user to export more information than just the localization of pathology. The pathology points are always part of the initial point cloud and to ensure that the dimensions, area and volume calculations will not suffer lack of information, both of the datasets (pathology and no-pathology) should be considered as parts of a common entity with different properties.

Future work will focus on the pathology extraction from more complicated surface of buildings' masonry than planar. The aim is that the users will be able to localize pathology and acquire as automatically as possible information on the dimensions of the cracks and fissures under their minor supervision.

REFERENCES

[1] M. L. V. Baptista, "Documenting a complex modern heritage building using multi image close range photogrammetry and 3D laser scanned point clouds", XXIV Int. CIPA Symposium, Strasbourg, France, 2013, pp. 675-678.

[2] S. Del Pozo, J. Herrero-Pascual, B. Felipe-García, D. Hernández-López, P. Rodríguez-Gonzálvez, D. González-Aguilera, "Multi-sensor radiometric study to detect pathologies in historical buildings", The Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Avila, Spain, 2015, pp. 193-200.

[3] J. García-Talegón, S. Calabrés, J. Fernández-Lozano, A.C. Iñigo, H. Herrero-Fernández, B. Arias-Pérez, D. González-Aguilera, "Assesing pathologies on Villamayor stone (Salamanca, Spain) by terrestrial laser scanner intensity data, Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Avila, Spain, 2015, pp. 446-451.

[4] J.I. SanJosé, J.J.Fernández-Martín, J.D. Pérez-Moneo, J. Finat, J. Martínez-Rubio, "Evaluation of structural damages from 3D laser scans", XXI CIPA Int. Symp.: Anticipating the Future of the Cultural, Athens, Greece, 2007.

[5] J. Shlens. "A tutorial on principal component analysis". arXiv preprint arXiv:1404.1100, 2014.

[6] M. Corral, "Vector Calculus", Michael Corral,2008.