

GENERATION OF GIGAPIXEL ORTHOPHOTO FOR THE MAINTENANCE OF COMPLEX BUILDINGS. CHALLENGES AND LESSON LEARNT

L. Perfetti^{a*}, F. Fassi^a, H. Gulsan^b

^a 3D Survey Group, Politecnico di Milano, Dept. of Architecture, Built environment and Construction engineering (ABC), Italy
(luca.perfetti, francesco.fassi)@polimi.it

^b hazalgulsan@gmail.com

Commission II

KEY WORDS: Photogrammetry, orthophoto, architecture, modelling, Cultural Heritage, 3D survey

ABSTRACT:

This study is part of the "Milan Cathedral Survey project". It is a three years long research project with the aim of surveying the entire cathedral in 3D with different techniques (mainly photogrammetry and laser scanning). The goal is to renew the architectonic drawings (sections elevations and plans) to obtain new updated and certificated measurements of the cathedral, as requested by Veneranda Fabbrica del Duomo, and to produce the basis on which to build a future 3D BIM (Building Information Model) system implementation. In this paper, we would like to examine in depth the survey process of the exterior elevations of the cathedral carried out using photogrammetry as the main survey technique for orthophotos production. The case studies here presented have the goal of underlying challenges, discussing decisions and approaches, describing the followed pipeline and of defining a standard method that can be followed to produce gigapixel orthophotos of complex cultural heritages.

1. INTRODUCTION

Nowadays in the research community operating in the field of new constructions but also in the field of Cultural Heritage (CH) the main research topic is BIM or HBIM (Heritage BIM). As well known, a BIM System can be generally and synthetically described as a "3D system" composed by one or more 3D models (parametric or semantic) of the building at different level of details that are connected to a more or less complex information system able to catalogue all the information regarding the building and its components, their interactions and the list of the maintenance activity carried out on over time (Oreni et al., 2017). This is a short and synthetic definition of BIM, which certainly does not want to detract from the importance of advanced building management systems. Researchers and private companies are developing BIM methods and tools increasingly efficient also in the field of CH (Chiabrando et al., 2017). It is clear, also in this field, the need for this type of systems precisely because they could be very useful tools for the maintenance of buildings over time. This has been tested during previous research projects conducted right on the Milan's cathedral (Tommasi et al., 2016; Fassi et al., 2015).

However, at present day, the complexity of the CHs, and the necessity for a deep geometrical and qualitative knowledge of the object does not go hand in hand with the degree of modelling simplification to which BIM systems force us today (Quattrini et al., 2015), making these systems of little use in everyday conservative practice. Indeed, inside the true reality of day by day maintenance and restoration projects, operators need big representation scale: 1:20, 1:50, accurate but synthetic representations of the buildings. Their activities are conducted elementwise on delimited areas that require precise interventions. Still today, geometric and measurable classical 2D representations remain the most used tool both in the operational phase and in the decision-making and design phase of all the activities.

In the near future BIM systems will forcefully become part of the conservation project pipeline of CHs, providing a context and a

global understanding of the building; but hardly will they replace the traditional ways of representing information for large-scale and focused interventions.

The project described in this paper aims at the creation of a completely new and updated "technical drawing package" that wants to be the "ultimate digital high accuracy" representation package of the whole cathedral.

1.1 The survey of the cathedral: external orthophotos.

The production of the whole drawing package requires a complete 3D survey of the cathedral, internal noble spaces, facades, roofs and also service spaces (Mandelli et al., 2017; Perfetti et al., 2017). The goal of the survey is to obtain a 3D point cloud of the whole cathedral with a minimum resolution of 5mm. This allows to extract 1:50 2D drawings (plotting error of 1 cm) when requested and to clearly detect the marble blocks subdivision of the wall structures that is one of the most important information for the conservation practices of the Veneranda Fabbrica (Fassi et al., 2015).

The idea was to use different 3D survey strategies to solve the various problems due to different "environmental conditions": the complexity of the architecture and the time dictated the choice.

For the external facades, photogrammetry was chosen as the primary survey technique. Laser scanning was mainly used for the interiors of the cathedral (to overcome illumination problems), it cannot be easily employed for the exteriors because the architecture develops itself too much vertically, leaving little space for manoeuvring on the horizontal plane, thus preventing a correct survey of the highest parts. The "bottom-up" acquisition results in too narrow scanning angles that could produce a non-uniform resolution, extreme edge effect and large shadow areas in the final point cloud. It would have made impossible to interpret the data correctly during the drawing phase and future 3D modelling.

Photogrammetry resulted to be the best solution for its positive aspects like high precision outputs and the extreme flexibility in

* Corresponding author

the survey. That means to have the possibility to reach skyward points of acquisition from which it was possible to survey the highest parts of the architecture, the horizontal and vertical parts at the same time, with uniform GSD (Ground Sampling Distance) and a correct acquisition geometry.

The original goal was to produce 2D drawings of the elevations of each facade complete with all decorations, spires, pinnacles, and statues with the primary purposes to draw clearly every single marble block composing the structure. Moreover, this survey was also needed to reconstruct the exterior 3D geometry of the church to complete and integrate the plans and the sections commissioned by the Veneranda Fabbrica. However, the technique choice also affected the final product: the exterior elevation drawings were substituted by digital orthophotos that are effective products to represent the marble block structure and to show their conservation status as well, giving additional information over the "health" of the surfaces.

So, the goal was triple: i) to create a unique gigapixel orthoimage of each facade at 1:50 representation scale, ii) to create 1:20, 1:10 orthoimages of every architectonic part and to build in this way the complete metric-photographic catalogue of every piece of the cathedral and iii) to extract horizontal and vertical profiles.

1.2 The orthophoto challenge

The orthophoto in architectural representation is a well-known method that originates from cartographic representation techniques: ortho-photography is the alternative to line mapping (Skarlatos, 1999) allowing at the same time the metric measurable representation of the geometry of the object as well as its quality/physical aspects. The history of ortho-photogrammetry is complex and has evolved rapidly only in recent years due to developments in computing power and technology. (Fassi et al., 2017). Till some years ago, it was used to survey and represent only "almost flat" or very simple façades, the so-called 2.5D objects, using image rectification processes. However, architectural or archaeological objects are usually much more complex, in particular at large scale of representation. The surfaces are discontinuous, very rich of details and in many cases, like the Milan's cathedral facades, each architectonic element is fully 3D. In this case, the ortho-projection should be more sophisticated because a simple rectified image would be metrically inaccurate.

In order to build a correct real-orthophoto, different projection planes (Pallaske et al., 1992), "break-lines", (Boccardo et al., 2009) hidden areas and 3D details must be considered. It means that it is necessary to use a real high-resolution 3D model of the object as a base for the ortho-projection. This is the reason why real-orthophoto projection was quite impossible until recently, creating a dense complete DSM (Digital Surface Model) of complex and large object is a very computational-demanding task.

For some years, laser scanning methods allowed to overcome the problem: very dense laser scans could be used directly to extract projected orthoimages (Georgopoulos et al., 2006), or could be used as a base DSM inside a co-georeferenced photogrammetric process.

Nowadays, photogrammetric image matching algorithms theoretically overcome the problem by offering inside the photogrammetric pipeline the possibility to build a very dense DSM directly from the oriented images (Deseilligny et al., 2011). This, in the photogrammetric community, is already a consolidated reality. However, from a practical point of view, it holds only for small and/or not very complex objects and remains challenging for very complex and extensive objects when the requested resolution is very high.

The presented project stresses the methodology investigating the possibility to create a single high-resolution (<5mm) orthophoto to represent very complex and extensive facades.

1.3 Three facades and one spire

In the paper four case study are described, three are the production of the gigapixel orthophoto of three facades of the Milan's cathedral and one is the production of the orthophotos of each side of the Amadeo's spire located on the rooftop. The first three cases of study investigate the possible methods to produce orthoimages of extremely large objects; the last one investigates the correct methods to survey very highly decorated and complex object at very high resolution (1:10-1:20) characterized by narrow spaces, repetitive geometries and highly vertical trend.

The first case study is the survey of the East façade (the apse side); the key aspect of this example is that the photos were taken without the use of means of elevation, exploiting only near building's windows and roofs. In this scenario, it was impossible to follow the perfect capture geometry with many later problems in the modelling phase.

The second described case study is the South façade where we used a lifting platform to reach 70 meters in height necessarily have the correct top-down view. This is a much better condition, here it is described the chosen ideal capturing geometry, the use of 3 different cameras and lenses, and the constant illumination issues.

The third case study is the survey of the North façade. It is very similar to the previous one: the ideal capturing geometry employing the lifting platform was used. The case of study describes some improvements over the South façade approach and presents the ideal results with no illumination issues.

The last case study is the survey of the Amadeo spire. It is one of the four complex middle size spires located around the cathedral's lantern (Figure 2 top, on the North-East corner). For this type of architecture, the 1:50 representation scale used for all the elevations orthophotos is not enough to describe the richness of decorations and the little complex blocks composition of the Amadeo's spire. Due to the impossibility to acquire images around the entire spire using telephoto lenses, to produce the orthophotos of the eight sides, the only option was to work up close to the spire exploiting the scaffolding, employing a wide-angle lens for the exteriors and fisheye lenses for the interior spaces (Perfetti et al., 2018). The consequent problems are the orientation of a huge number of photos, the connection between different scaffoldings levels and the orientation of all the data together.

2. CHALLENGES AND REQUIREMENTS

Surveying a complex architecture with several geometrical attributes and details is a complex work, considering the size of the building, the difficulty of the job increases. The complexities of the building required a multi-scale survey to assure the desired precision (Fassi et al., 2011). All the surveys here described, therefore, were made by using multiple lenses, different capturing geometry and using external measurements with a total station for geo-referencing, accurate scaling and to check every survey step.

Many are the difficulties and the topic to consider both during the acquisition and during the processing:

2.1 Logistics

The Milan's cathedral, like many other iconic cultural heritages, as for example the Basilica di San Marco in Venice (Adami et al., 2018), is a hub of activities in which many actors play different interconnected roles. It is a place of worship with a tight

celebration schedule, a place to visit with lines of tourist every day and a valuable monument took care by a continuous restoration yard in charge of the maintenance. The survey activities must be carefully planned and organized to fit the cathedral schedule and at the same time, they must be flexible and resilient to sudden changes.

The survey activities of the South and North façade used the aid of a lifting platform of which rent could not be rescheduled due to unideal light conditions thus it was necessary to overcome by post-processing any unpleasant light variations (Figure 3).

2.2 Multi-lenses survey

The complexity of the case study in its architectonic configuration and elements require to employ more than one focal length throughout the survey. The three main planes of the facades, the street elevation, the first level of the roofs and second level of the roofs are many meters apart from each other and therefore, those areas will be rendered at very different resolutions when framed from the ideal plan of acquisition (a vertical plane in front of the façade). Focal length ranging from 24mm up to 85mm (on full frame sensors) were used for the survey of the south façade. Others specialised lenses were needed for the survey of the Amadeo spire, an 8mm fisheye and a 12mm rectilinear and a 105mm was used for some detail integration.

It follows that it is crucial to use a software that supports multiple focal lenses in the same project and can estimate internal parameters of all of them at once. Pre-calibration becomes very impractical considering that the focal plane also must be changed passing from acquisitions from the ground to acquisition on the roofs or from picture taker from a building to picture taken from a scaffolding.

2.3 Image resolution

The Veneranda Fabbrica of the Milan's cathedral requested the orthophotos of the exterior elevations and of all the roofs elements and surfaces, like the flying buttresses, at the scale 1:50, therefore the resolution was defined accordingly by playing with focal length and distance from the object. However, an important distinction must be made between the resolution of the pictures and the resolution of the point cloud/model in terms of features extraction. The former depends only on images resolution, and image quality from a photographic standpoint is an important metric; the latter instead, depends greatly from the intersection angles (Luhmann et al., 2013) that originates from the network geometry.

It is crucial to define the aim of the survey; indeed, the two approaches push toward two very different network design based on different priorities.

To produce high-resolution orthophotos, good quality high-resolution nadiral pictures are fundamental, while to produce high accuracy dense clouds/models for features extraction the network geometry becomes more important and must be strongly three-dimensional (convergent image pairs) around each significant architectonic element that is expected to be accurately modelled, like decorative pilasters. Indeed, provided that the images to be used for the orthophoto construction are nadiral to the object and parallel to the projection plane, taking it to the extreme, even a rough model can be sufficient to obtain a metric representation. Similar resolution images non-nadiral to the same rough model will instead result in a very distorted orthophoto. In the latter case, when forced to use only angled pictures, a very precise model is mandatory to obtain a metric result.

It follows that the time investment both in the acquisition campaign and in the elaboration phase to obtain a high accuracy

model (if possible at all), is only worth it in the lack of nadiral images.

The network geometry used in our tests described in section 3.1 were aimed at the construction of orthophotos only and was therefore significantly simpler than it would have been if intended for features extraction.

2.4 Capturing geometry

As mentioned above, the capturing geometry depends mainly on the goal of the survey, we chose to use a network geometry oriented towards orthophoto production and therefore it was based on nadiral images.

Being able to take pictures from a high level was mandatory, when possible we relied on the presence of existing buildings close to the cathedral, like the Veneranda Fabbrica's headquarters building from which windows it was possible to acquire all the pictures needed for the East façade elevation.

A track lifting platform (Figure 1) was used to reach the needed height for the acquisitions of the North and South facades.

The ideal capturing geometry we chose to use with the lifting platform was to go up vertically in front of each glass window of the cathedral rising 2/3m at a time, from each position we take three pictures: one nadiral and two tilted left and right to better survey the pillars and to avoid shadow areas around statues and decoration.

For the cathedral roofs top, a standard close network was followed for each of the roof's spans: nadiral images from the four sides pointing towards the opposite side plus some images to strengthen the connection in the corners (Figure 4).



Figure 1. views of the cathedral's exteriors: acquisitions from the platform for the South façade (top right) and for the North façade (top left). On the bottom a panoramic view of the South roofs from the platform. Notice the complexity of the decorative elements and the challenging illumination conditions.

2.5 Multi-block approach

The case studies are extremely complex and large in their extension that a subdivision in “blocks”, areas, is almost compulsory. For the construction of the façades, we identified enclosed architectural areas (“roof’s spans”) that could be processed together or independently depending on the process. The network alignment was segmented in large macro areas: the apse façade, the apse roofs area, the north façade, the north roof area, the south facade and the south roofs. the dense image matching process was performed on smaller blocks to contain the otherwise too large number of images to be handled at the same time, i.e. individual roof’s spans (Figure 2 top). Likewise, although the pipeline was not consistent from start to end, the blocks were further segmented for the model reconstruction process focusing on individual architectural elements i.e. the flying buttresses, the roof covers, the gothic embrasure etc. The same element subdivision was kept then for the orthophoto production and, only during the final mosaicking, all the orthophotos of the different parts were stitched together. Although the processing phases benefit from a block division, it is primarily important to ensure coordinates consistency over the process. Starting from the alignment phase, even though the facades and the roofs were processed separately, great attention was played to verify their geometric coherency. That was done by using non-coded circular targets that could be clearly framed both from the roofs and from the facades’ acquisitions using the lifting platform. Other coded targets were placed where possible (roof areas) to improve control and constraint. Both targets were measured and reference to a close topographic network that surrounds the cathedral. The non-coded targets were used to perform a rigid georeferencing of the macro blocks to check that deviations were contained under 1cm error. The common coordinates system allowed also to establish a common projection plane for each of the facades to be used for the orthophotos production. The generation of the georeferencing text file together with each of the individual orthophotos allowed for easy mosaicking at the end (Figure 2 bottom).

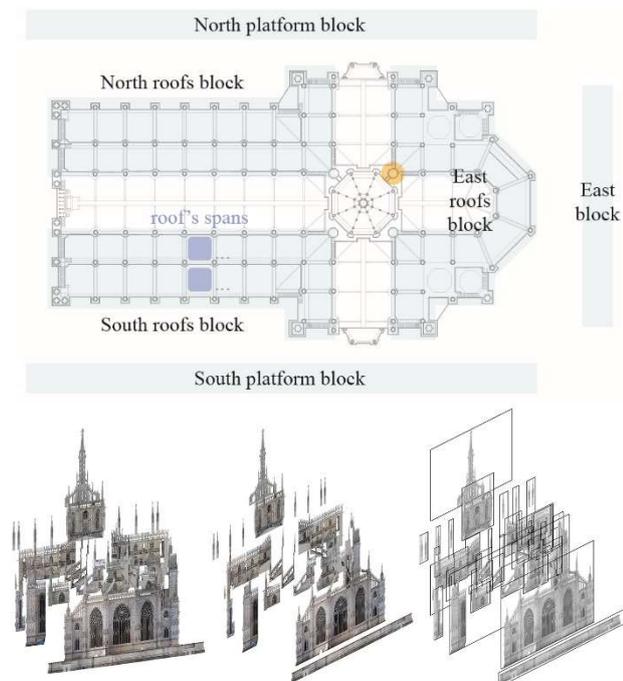


Figure 2. block division: alignment segmentation (light blue), dense matching segmentation (blue), Amadeo spire (yellow). Mosaicking of the orthoimages (bottom).

2.6 Constant illumination

One of the greatest challenges to this work was the illuminations. Constant illumination is the most important elements that control the visual appearance of orthophotos and allows a correct visualization of the marble blocks subdivision. The best approach is usually to perform the survey on a cloudy day to avoid sharp changes in light condition throughout the acquisition. However, due to logistic issues like scheduling long in advance the use of the lifting platform, that required to close portions of the square/streets, or the access to each window and rooftop of the Veneranda’s building; it was mandatory to define a strategy to overcome the shift in light conditions during the post-processing. We chose three different strategies depending on the scenario:

2.6.1 Repeat the acquisition: for the East façade, in front of the Veneranda Fabbrica’s building, we repeated portions of the first acquisition that was carried out on a sunny day. It was possible to manage to access for the second time only some of the windows giving very short notice. All the pictures, main acquisition and integration on a cloudy day, were used together for the model construction, but only the second ones were used for the orthophoto generation.

2.6.2 Integration with different setup: similarly, an integration of new pictures was done for the south façade. Due to the impossibility of repeating the survey with the lifting platform, the new pictures were taken only from the ground floor. An 85mm lens on a 36-megapixel camera was used to ensure a similar resolution on the top of the cathedral. The integration from the ground was only sufficient to solve the problem for the street elevation of the cathedral.

2.6.3 Manual HDR (High Dinamic Range) processing: the last strategy was used for all those areas that couldn’t be reached without the lifting platform. Some of the images taken from the platform (the strictly sufficient number to cover the façade) were processed using an image editing software. The image was copied 2 times and was overexposed and underexposed by 2 stops, manual masking was used to merge the three variations of the same image together to remove the shadows (Figure 3).

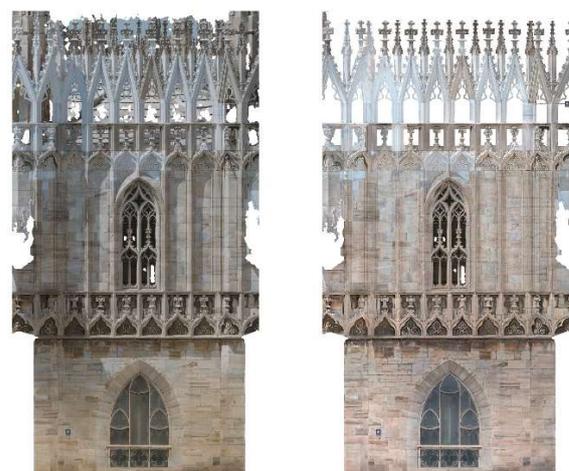


Figure 3. orthophoto of a portion of the South elevation, raw result on the left, and final on the right after the post processing.

2.7 Post processing:

To achieve the best result, post-processing is a mandatory phase, against what we expected this phase was very time consuming and – based on our needs – it cannot be avoided. This phase can

be divided into two parts: the processing of the images and the processing of the orthophotos. All the images were processed to get the most consistent look possible, although precise colour correction was not required, all the images were manually adjusted using the software Adobe Lightroom. A second adjustment was performed on the final orthophotos to make sure they would match on the mosaic.

However, the most time-consuming post process was by far the contour cleaning of each of the orthophoto to remove incorrect data on the edges of very complex decorative elements. Because the 3D model of the façade is never complete but always one-sided, the contour of highly decorated elements gets rendered jagged. We manually traced the edge of each orthophoto to delete the incorrect data.

The contour cleaning was indispensable to ensure the best result in terms of the final look of the mosaic. The manual cleaning allows obtaining sharp edges for the decorative gothic embrasure that better describe the element itself and allows to see-through what is behind (Figure 6).

Although the raw orthophotos contour would improve by improving the accuracy of the model, a balance decision should be made between time spent on the elaboration phase and time spent on the post-processing phase.

3. THE CASE STUDIES

3.1 The survey strategy

3.1.1 East façade: The survey of the East façade was the first one to be carried out and it relied mainly on the presence of the Veneranda Fabbrica's building located just in front, the windows rhythm dictated the base distance between the acquisitions on the horizontal plane as well as on the vertical one.

The acquisition process can be divided into three main groups: i) the pictures taken from the ground, ii) the picture taken from the Veneranda Fabbrica's headquarters and iii) the pictures taken on the roof's level. All the pictures were taken using the full frame DSLR Nikon D810, a 50mm lens was used for group 1 and 2 and a 24mm for group number 3. From the ground, the pictures were framed in portrait and the acquisition followed arches around each corner of the apse basement plan (Figure 4 left). from the building the pictures were shot horizontally, five per window, one nadiral, two tilted left and right and two tilted up and down. A second acquisition was carried out for group 1 and 2 to integrate few pictures under cloudy weather conditions. Pictures of group 3 were acquired immediately on a cloudy day due to much simpler scheduling of operations. For each roof's room, a classical network geometry was employed (Figure 4 right).

For group 1 and 2, the GSD on the street elevation was less than 5mm, for group 3 it was 2mm.

The described acquisition produced good results in terms of metric accuracy of the reconstruction of the main elevation and in terms of visual appearance, especially thanks to the later image integration. However, the upper parts of the façade show some problem because all images were acquired with a bottom-up point of view. The maximum height of the Veneranda Fabbrica's building forced a limit on the resolution of the 3D model obtainable from the image acquisition network. Specifically, the highest level of the gothic embrasures was only covered with tilted pictures taken from the rooftop of the building pointing upwards and therefore the resulting model suffered from distortions and stretches on the upper parts of all the decorative elements that could be solved only afterwards by adding extra pictures taken from the lifting platform. The resulted orthophotos of those parts were of poor quality. The spires will have to be surveyed again using a different approach (drone acquisition), while the gothic embrasure area was constructed correctly after

the integration from the platform. Figure 8 shows the final product.

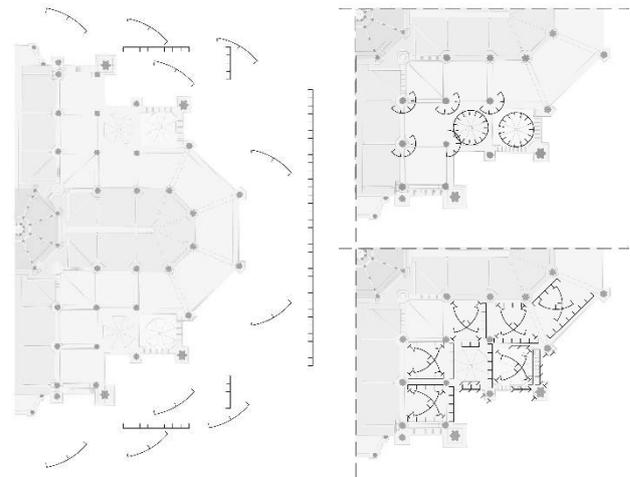


Figure 4. network geometry used to survey the apse, on the left the network followed on the ground, and on the right the network of the roofs.

3.1.2 South façade: The acquisition can again be divided into the same three groups as the East façade. However, the capturing geometry was drastically different. For the South façade, a track lifting platform was rented from the very beginning to avoid the problems that rose up in East façade. From the platform, three set up were used:

- *Nikon D810 with 50mm lens:* it was used to take the pictures from a height of 5 meters to the top, 70m. the lens was chosen to ensure a GSD of 4mm on the elevations of the first and second level of the roofs. The GSD on the street elevation was around 1.5mm resulting in a reconstruction suitable for the scale 1:20.
- *Canon 5D markIII with 35mm lens:* it was used to take the pictures of the street elevation only, up to a height of around 30m, the resulted GSD was of 1.8mm, perfectly suitable for the scale 1:50. For the bottom part, the pictures taken with this configuration can be used in complete substitution to the pictures taken with the previous configuration.
- *Canon 5D markIII with 85mm lens:* it was used from a height of around 50m up to the top. It allowed a GSD of 2.4mm on the upper level of spires and allowed to reconstruct the orthophoto of those elements. The same camera configuration was used to integrate a few images from the ground on a cloudy day as mentioned in Section 2.6.2.

Six pictures were taken at each position of the platform, from the bottom up to a height of approximately 30 metres, three with the 50mm and three with the 35mm in a configuration of nadiral plus left and right images. From 30m upward the same approach was carried on with the 50mm, while with the 85mm, starting from approximately 50m, a picture for each spire was taken in a radius of about 3-4 spans.

The same set up and capturing geometry of the East façade was kept for the roof level: the Nikon D810 with the 24mm. A total of 7586 pictures were used to complete the reconstruction of the South façade (Figure 9).

3.1.3 North façade: The North façade acquisition process resembles the South one almost entirely. Fewer pictures were acquired this time from the lifting platform, mainly by reducing the picture taken with the 50mm under the 30m range and by reducing the number of pictures taken with the 85mm as well. On the contrary, more pictures were acquired for the acquisition from the roof's to better survey architectonic detail in the roof's tiles and flying buttress. In this area, the illumination conditions were ideal, all the architectonic elements are exposed to the north, so everything is in shadow under uniform illuminations, and there was no need for later integration under different weather. Figure 10 shows the result.

3.1.4 Amadeo spire: The last case study is the survey of the Amadeo spire, a very richly decorated spire located on the North-East corner of the lantern. While it was possible to survey the spires with the aid of the lifting platform and the possibility to go up to 70 metres height, only one side, the street side of the spires could be surveyed. In order to survey all eight sides of the Amadeo spire, and to do that for the scale 1:20, as asked by the Veneranda Fabbrica it was requested to change approach completely, further increasing the challenge of a multi-scale, multi-block and multi-lens survey described in Section 2.

As it was not possible, to carry out the survey using drones, we opted for a classic close-range network. It wasn't possible to survey all the decorative elements from the rooftop only at the scale 1:20, therefore, we were forced to wait for the installation of the scaffoldings around the spire for the restoration activities. On one hand, the presence of the scaffoldings allowed us to acquire nadiral images up to the very top of the spire, on the other hand, it places the challenge of connecting the different levels together and of taking images very close to the object.

The camera configuration used was the Nikon D810 with a 12mm focal length, it allowed at the same time to maximise the overlapping between neighbouring images and to connect the scaffolding elements via small openings close to the spire to be framed very up close from the bottom and from the top of each level. The GSD was of 0.8mm, sufficient for scale 1:10.

The use of coded target was mandatory to check the correctness of the alignment with the aid of topographic measurements.

The inside part of the Amadeo's enclosed spiral staircase was also surveyed for the orthophoto production of the eight sides of the central pillar. A preliminary description of this work, carried out using fisheye lenses, can be found in (Perfetti et. al. 2018). Figure 11 shows the results.

3.2 The photogrammetric proposed pipeline

All the images were acquired in raw file and processed to adjust brightness level and colour balance before the photogrammetric process. This phase as to be considered optional, especially considering that shooting in raw, although improves the versatility of the images allowing strong adjustment, it also increments evilly the data size.

After the acquisition phase, the images were processed following a classical photogrammetric pipeline (especially for the alignment phase) using the software Agisoft Photoscan. (Nocerino et al., 2014; Fassi et al., 2015). Different approaches and attempts were instead carried out regarding the strategy to build the DSM (matching + mesh phase). Figure 5 synthesizes the pipeline we found to be ideal, and Table 1 gives an idea of the subdivision of the efforts for each phase.

The alignment process was performed at maximum resolution (accuracy settings = high) subdividing all the object into macro areas of thousands of images (Figure 2 top) so that the GSD of the images was more or less uniform within the photogrammetric blocks. In this way, for example for the South façade, the pictures

were divided into two groups, i) middle-range pictures from the lifting platform and ii) close-range pictures from the roofs. All the orientation sets were then checked and optimized using the targets (coded and non-coded) that were measured with the total station.

The dense image matching was performed using "high" as accuracy settings dividing the process into smaller groups of around 100 images. This was necessary to keep the data manageable and speed up the matching process.

The software CloudCompare was used to subsample the point clouds prior to the mesh generation in order to ensure a uniform resolution of the point cloud of 4mm. The decimated point clouds were reimported in Photoscan and the chunks were further subdivided into architectural elements for the mesh model production.

Different approaches have been followed for the mesh generation during the entire work, normally the blocks were processed using the full point cloud resolution mostly without any interpolation. The interpolation, as intended by Agisoft PhotoScan, was used for border areas, rich of decoration for which there could be small gaps in the point cloud.

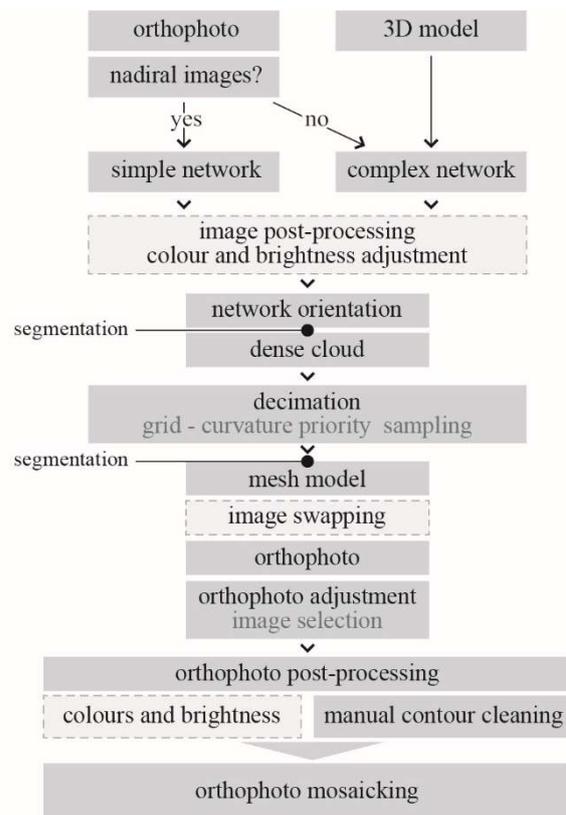


Figure 5. Pipeline for orthophotos production.

phase	time required
acquisition phase	5%
images post processing	2%
process till point cloud	30%
decimation	3%
process till orthophoto	20%
orthophoto post process	30%
mosaicking	10%

Table 1. Time demand of the different phases considering a total processing time of about 6 months with 1 operator.

At this point, it was possible to substitute the images used in the process so far with more strongly processed versions to improve, if necessary, the orthophoto result. For the South façade, some images were substituted with HDR versions.

The orthophotos were produced and adjusted using Photoscan. The possibility to select the best images to project on the model and to modify by hand the orthophoto seamlines that the software offers has been especially important to produce orthophotos in presence of obstacles like the scaffoldings.

At the end, the orthophotos were post-processed using an image editing software, i.e. Adobe Photoshop to manually trace and clean the contour (Figure 6) before they could be mosaicked together (Figure 2 bottom) to complete the façade (Figure 8).

4. LESSON LEARNT AND DISCUSSION

It is certain that to produce such complex orthophotos requires a lot of effort, it is a long process and it is costly. It is therefore imperative to have the aim of the work clear and to distribute the efforts in an effective way. Although the metric accuracy of the final orthophoto must be guaranteed for the expected representation scale, to exceed it can be very expensive in terms of computational efforts, time and costs. The acquired experience, from the first case (east façade) to the last one (north façade) led to finding the correct balance between a sufficient level of completeness and accuracy of the DSM and final ortho-mosaic result. Here some consideration about choices and reached compromises

4.1 Nadiral images vs accurate model

How does image resolution and model resolution relate to one another in orthophoto-oriented network design? Both are fundamental and the second depends on the first one. Image resolution must at least match the representation plotting error following the well-known relation: $GSD \leq \text{plotting error}$; model resolution is one of the main factors that ensure metric accuracy. However, if the aim of the photogrammetric acquisition and process is the production of orthophotos only, we can state that the resolution on the x-y plane, the projection plane, is much more important of the resolution on the z-axis. That is what naturally occur out of a network design as the one we used, with a plane of acquisition, where the images were taken, ideally parallel to the plane of projection (Luhman et al., 2013).

We repeatedly observe that for a network geometry – such that images planes are not parallel to the projection plane – to produce accurate orthophotos there must be a very precise model “behind”; whereas a sufficient model (very precise on the x-y plane even if poor on the z-axis) is enough if coupled with nadiral images. Figure 7 (top image) shows a slice of the point cloud generated from the East façade acquisition, the profile of the pilaster is of very poor resolution, it deviates significantly from the real one, more than the tolerance of the 1:50 scale. However, in the resulting orthophoto, because of the use of nadiral images, no distortions could be measured.

A network geometry designed for orthophotos is simpler and require less image, moreover, it speeds up the processing phase and is less prone to error. It follows that provided that is possible to acquire nadiral images, to our experience, such a network geometry is favourable to more complex others.

The first and most important condition to check that determines together the quality and the required time of an orthophoto is the accessibility, whether is possible to acquire nadiral images and therefore keep the network simple or not.

The spires of the East façade couldn't be effectively surveyed from the Veneranda's headquarters rooftop despite the GSD being sufficient.

4.2 Metric accuracy and visual appeal

As anticipated in Section 3.2, for the highly decorated elements, like the gothic embrasures, the biggest issue to solve was about the 3D model: it was preferable to get a rough but “generous” mesh model (Figure 6) then a more exact but still jagged one in particular on the edges. The answer comes back to effort management again. Two where the reasonable approaches: i) to mask all of the images to obtain a cleaner dense cloud that will not require a manual post-processing on the final orthophoto; or, on the contrary ii) to not mask the single images, getting a rough and noisy point cloud on the edges that results in a mesh with “abundant” and jagged edges and, subsequently, deal with the final orthophoto by manually cleaning borders and masking unwanted and false projections.

However, such complex elements would require very elaborated and heavy networks to reconstruct an extremely precise geometry. Aiming at producing orthophotos only, and therefore relying on nadiral images mainly, the edges of the model would result jagged anyway, even as a result of masked images. Moreover, because of the limited geometric resolution of the mesh model, an exact border, could as well result in losing parts of the actual architectonic element.

The rough edges of the mesh model could indeed “save” parts of the geometry that would otherwise be lost by masking. And the time required to manually clean the contour of one orthophoto, at this point compulsory, is less than the time require to mask all the images that produce it.

The suggested approach, chosen and followed in the last case of study, was to not edit or adjust the raw mesh model, and instead to clean the ortho-mosaics afterwards. The visual appearance of the manual cleaned orthophotos outmatch the raw results obtainable from masked images (Figure 6). Even though the orthophoto post processing is one of the more time-consuming phases of the process, it saves time in comparison to the alternatives.

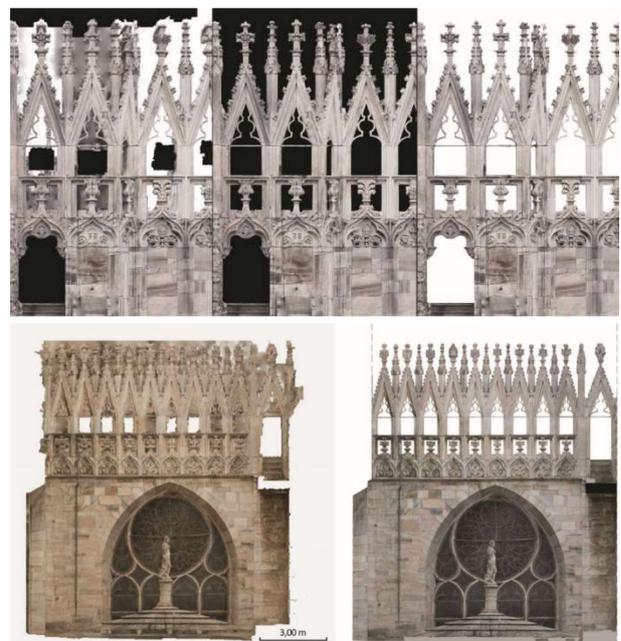


Figure 6. post processing of the orthophotos, manual masking of the rough edges resulting from an “abundant” mesh model that exceed the actual edge of the gothic embrasures.

5. CONCLUSIONS

During the experience with the Milan's cathedral's orthophoto production, many different approaches and strategies were tested facing many challenges that are typical of this kind of projects, like the illumination issue, the logistic, the weight of the data, the integration of different software and so forth. The definition of the aim of the survey and means of accessibility strongly defines the network geometry and complexity and cascading determines the computational demand of the processing.

One of the most tested aspects is the level of accuracy and resolution that the DEM (Digital Elevation Model) should have to ensure an accurate, complete and also "beautiful" orthophoto. The DEM resolution is an important aspect, however, from the point of view of the production of ortho-mosaics, it passed almost unnoticed while many other factors like the light and the resolution coherency were much more important. Nadiral images captured by a correct elevated position are "necessary and sufficient condition" for a correct orthophoto. Images that were taken under strong constraints (as for the first east facade attempts, later updated) could not guarantee correct results even after a long manual correction and post-process.

The creation of the DEM as described, in terms of resolution and capture geometry, worked well for orthophoto production but not for 2D features extraction for plans and sections drawings (Figure 7, top image). For this purpose, the second aim of the survey, the resolution should be increased, and the capture geometry reinvented to acquire the complex space with a more complete three-dimensional vision. A challenging and time-consuming job that requires maybe more complex matching algorithms able to work in reasonable time with full resolution images. In order to overcome the problem, additional localized micro-survey at very high resolution were performed in order to extract profiles, and details necessary to ensure 1:20-1:50 drawing resolution (figure 7 centre).

The orthophoto of the Milan Cathedral is now used by Veneranda Fabbrica to set up a sort of GIS (Geographic Information System) related to the state of health of the structures in order to monitor the level of danger and risk of the building.



Figure 7. Plotting of the photogrammetric (red) and laser scanning (yellow) dense cloud (top image). High resolution photogrammetric profile obtained by later integration (centre). The point cloud used for the plan restitution (bottom).

ACKNOWLEDGEMENTS

Veneranda Fabbrica of the Milan's Cathedral with the director Eng. Francesco Canali and Francesco Aquilano and Massimiliano Regis supported this work. The authors sincerely thank all those (students and trainees) who collaborated on this project. A special thanks to Andrea de Biasi, involved in the project for his master thesis together with Hazal Gulsan, co-author of the presented paper and Carlo Polari that participated to the survey.

REFERENCES

- Adami, A., Fassi, F., Fregonese, L., and Piana, M., 2018. Image-based techniques for the survey of mosaics in the St Mark's Basilica in Venice. *Virtual Archaeology Review*, 9(19), 1-20. doi:10.4995/var.2018.9087.
- Boccardo, P., Dequal, S., Lingua, A. and Rinaudo, F., 2009. True digital orthophoto for architectural and archaeological applications. *Geospatial world*. <https://www.geospatialworld.net/article/true-digital-orthophoto-for-architectural-and-archaeological-applications/> 2018.
- Chiabrando, F., Lo Turco, M. and Rinaudo, F., 2017. Modeling the decay in an HBIM starting from 3D point clouds. a followed approach for cultural heritage knowledge. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W5, 605-612, doi:10.5194/isprs-archives-XLII-2-W5-605-2017.
- Deseilligny, M., De Luca, L. and Remondino, F., 2011. Automated Image-Based Procedures for Accurate Artifacts 3D Modeling and Orthoimage. *Geoinformatics FCE CTU*. 6, pp. 291-299.
- Fassi, F., Achille, C., Mandelli, A., Rechichi, F. and Parri, S., 2015. A new idea of BIM system for visualization, web sharing and using huge complex 3D models for facility management. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-5/W4, 359-366, doi:10.5194/isprsarchives-XL-5-W4-359-2015.
- Fassi, F. and Campanella, C., 2017. From daguerreotypes to digital automatic photogrammetry. applications and limits for the built heritage project, In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-5/W1, 313-319, doi:10.5194/isprs-archives-XLII-5-W1-313-2017.
- Fassi, F., Achille, C. and Fregonese, L., 2011. Surveying and modelling the main spire of Milan Cathedral using multiple data sources. *The Photogrammetric Record*, pp. 462-487.
- Fassi, F., Rossi, C. and Mandelli, A., 2015. Emergency survey of remote and endangered archaeological sites. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-5/W4, 85-91, doi:10.5194/isprsarchives-XL-5-W4-85-2015.
- Georgopoulos, A. and Natsis, S., 2006. A simpler method for large scale digital orthophoto production. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVII. Part B5.
- Luhmann, T., Robson, S., Kyle, S., and Boehm, J., 2013. *Close-Range Photogrammetry and 3D Imaging*. De Gruyter, Incorporated.

Mandelli, A., Fassi, F., Perfetti, L. and Polari, C., 2017. Testing Different Survey Techniques to Model Architectonic Narrow Spaces. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W5, pp. 505–11, doi:10.5194/isprs-archives-xlii-2-w5-505-2017.

Nocerino, E., Menna, F. and Remondino, F., 2014. Accuracy of typical photogrammetric networks in cultural heritage 3D modeling projects. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-5, 465-472, doi:10.5194/isprsarchives-XL-5-465-2014.

Oreni, D., Brumana, R., Della Torre, S. and Banfi, F., 2017. Survey, HBIM and conservation plan of a monumental building damaged by earthquake. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-5/W1, 337-342, doi:10.5194/isprs-archives-XLII-5-W1-337-2017.

Pallaske, R., Marten, W. and Mauelshagen, L., 1992. Digital Orthophoto System for Architecture Representation. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XVIIth ISPRS Congress Technical Commission V: Close-Range Photogrammetry and Machine Vision. August 2-14, 1992, Washington, D.C., USA. pp 67-73.

Perfetti, L., Polari, C. and Fassi, F., 2018. Fisheye multi-camera system calibration for surveying narrow and complex architectures. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2, 877-883, doi:10.5194/isprs-archives-XLII-2-877-2018.

Perfetti, L., Polari, C. and Fassi, F., 2017. Fisheye photogrammetry: tests and methodologies for the survey of narrow spaces. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W3, pp. 573-580, doi:10.5194/isprs-archives-XLII-2-W3-573-2017.

Quattrini, R., Malinverni, E. S., Clini, P., Nespeca, R., and Orlietti, E., 2015. From TLS to HBIM. high quality semantically-aware 3D modeling of complex architecture, In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-5/W4, 367-374, doi:10.5194/isprsarchives-XL-5-W4-367-2015.

Skarlatos D., 1999. Orthophotograph production in urban areas. *Photogrammetric Record*, 16(94), pp. 643–650.

Tommasi, C., Achille, C. and Fassi, F., 2016. From point cloud to BIM: a modelling challenge in the cultural heritage field. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B5, 429-436, doi:10.5194/isprs-archives-XLI-B5-429-2016.



Figure 8. Orthophoto of the East façade: 3580 images, around 1.2 billion pixels (GSD 2mm).



Figure 9. Ortophoto of the South façade: 7586 images, around 1.9 billion pixels (GSD 2mm), 156m long.

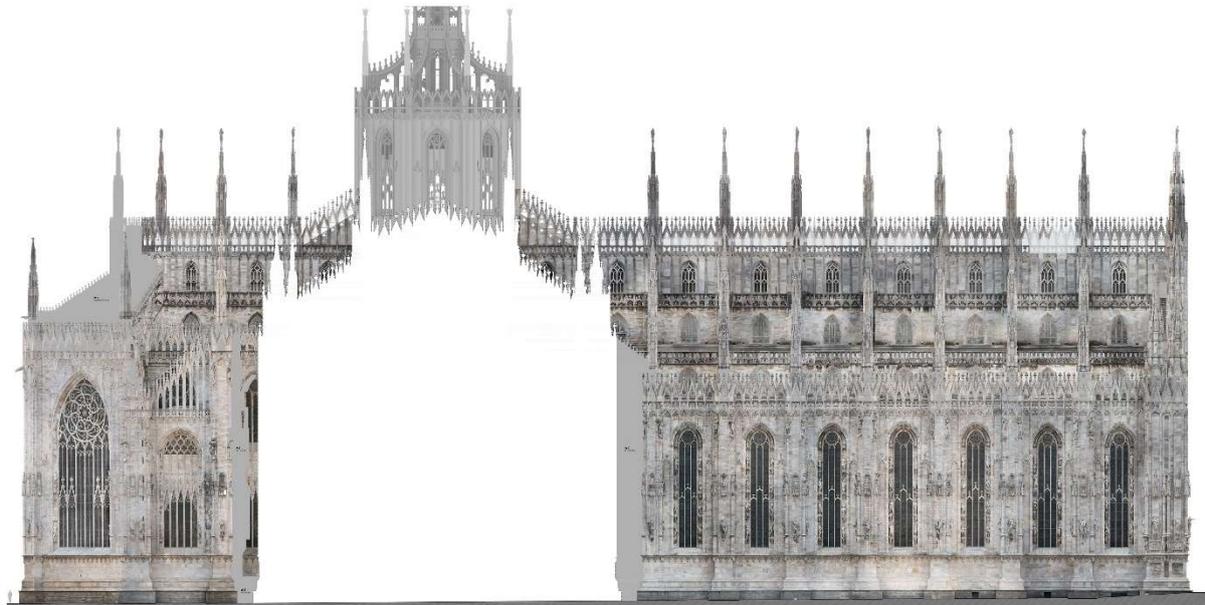


Figure 10. Ortophoto of the North façade: 4902 images, around 1.4 billion pixels (GSD 2mm), 156m long.

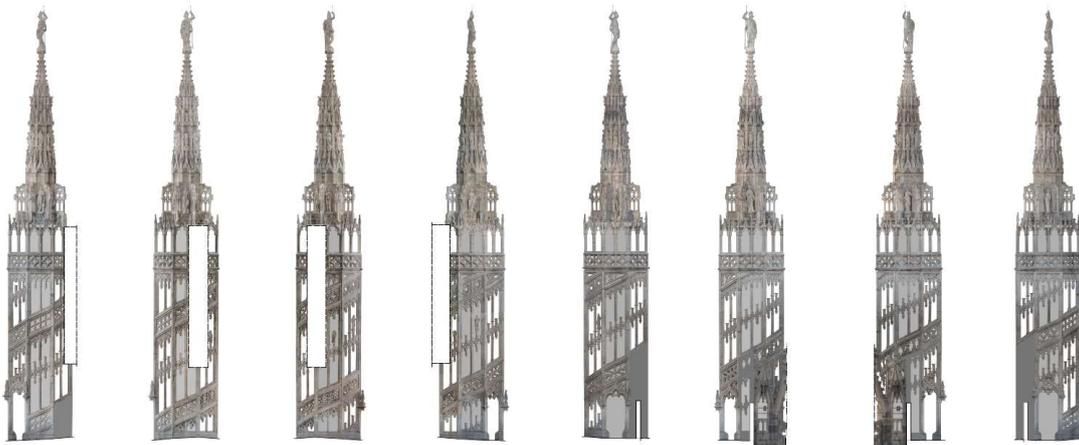


Figure 11. The 8 orthophotos of the sides of the Amadeo's spires at 1:10 representation.