THE APPLICATION OF CONTENT-BASED INDEXING AND RETRIEVAL OF DIGITAL PICTURES AND RANGE IMAGES TO THE 3D VIRTUAL RECONSTRUCTION OF COMPLEX SITES AND ARCHITECTURES

E. Paquet [a*], J.-A. Beraldin[a], H. L. Viktor[b] and Benedetto Benedetti[c]

[a] Visual Information Technology, National Research Council, M50 Montreal Road, Ottawa, K1A 0R6, Canada –
[b] SITE, University of Ottawa, 800 King Edward Ave., Ottawa, K1N 6N5, Canada – hlviktor@site.uottawa.ca
[c] Scuola Normale Superiore di Pisa, Palazzo del Capitano, Piazza San Felice, 7, Pisa, Italy - b.benedetti@sns.it

Commission VI, WG V/2[a]

KEY WORDS: Content-based indexing/retrieval, model, range images, registration, three-dimensional, 3D

ABSTRACT:

The modelling of complex objects and sites involve the acquisition of a large number of texture maps and range images; each one of them representing a particular viewing angle. These views must be combined and registered in order to create an accurate model of the original. The complexity of the resultant models, and consequently the number of views required, has increased tremendously over the past decade. Furthermore many major projects involve multinational and multilingual teams, each with different underlying methodologies. In addition, different teams may survey the same object or site over a number of years, producing complimentary results on different platforms utilizing a variety of (possible outdated) technologies. In such conditions, it is difficult to make sense of the annotation and to determine which views should be registered together. We propose a new approach in which similar views are found by content-based indexing and retrieval of textures (2D) and range images (3D). The views are described automatically, according to their appearance and their 3D geometry. A search engine allows retrieving similar and related views for registration.

1. INTRODUCTION

Complex object and site modelling both involve the acquisition of numerous texture and range images. Each range image (3D images) is associated with a certain partial view. These partial views must subsequently be registered and merged with one another, in order to form a coherent and accurate model of the object or site under consideration. In recent years, the number and the complexity of such views produced with 3D imaging systems have been increasing tremendously. The same can be said about the projects themselves. Most projects involve more than one team and 3D imaging system. Very often, the teams differ by their country, their language (e.g. English, French, or Japanese) and the methodology by which they annotate the views (sequential, but meaningless, naming of files, spelling errors, different languages, repetitive files names, etc.). Furthermore, there might be a large interval of time in between the different sets of acquisitions (legacy data). In these conditions, it becomes very difficult to efficiently perform the registration and the merging of the views. Thus, a substantial amount of time is devoted trying to make sense of the other group's methodology, if any. In the worst case, which is not uncommon in practice, the later might even have been discarded. Another situation might occur when a structure is physically taken apart for preservation or conservation work and it is required to reconstruct it virtually or physically. Proper documentation could be, in such a situation, sparse or even absent. Furthermore, the different pieces could have been spread around the world or simply have been mixed up. One solution can be to digitize, in 3D with texture, the different elements and then use our approach to assist in putting the pieces together for a virtual reconstruction. This phase can be followed by an actual reassembly of the original structure.

In order to address these problems, we propose a new approach based on content-based indexing and retrieval. In our method, the pictures and the views are analysed automatically, in order to generate a compact and abstract index that describe respectively the composition of the 2D image and the three-dimensional shape of the views. By image, we mean pictures acquired with a digital camera or multi-spectral imaging device, or colour intensity images acquired with a laser scanner. The ensemble of all these indexes therefore constitutes a database that can be searched with a dedicated search engine. The search engine is based on the query by example paradigm: the user chooses an image or a view and the search engine retrieves the closest ones from the composition or 3D shape point of view. In this way, the different views created by diverse team can rapidly and efficiently be registered and merged, irrespectively of their origin or the underlying methodology.

Our paper is organised as follows. Firstly, we present the general framework of acquisition and modelling with laser scanners and digital cameras. Then, we investigate the issues involve in large and multifaceted projects related to the virtualisation of complex object by multinational teams over a long period of time. We subsequently describe the content-based aided approach designed to solve these problems. We describe our indexing methods, both for 2D images and range images, as well as our query by example system. Finally, we apply our method to real problems of acquisition and modelling and present our conclusions. This paper is an extended version of the one presented by (Paquet and Beraldin, 2007).

2. REGISTRATION OF RANGE IMAGES: AN OVERVIEW

Because of object or site dimensions, surface topology complexity, occlusions, and required spatial resolution, it is

* eric.paquet@nrc-cnrc.ca.ca

usually necessary to acquire multiple 3D images and texture images from different view points, in order to cover every surface at the desired level of details. These 3D imaging systems generate on average between 10 000 to 500 000 3D coordinate per second (Blais 2004). Yet, the acquisition of 3D images is considered a time consuming step because of the omnipresent next best view problem. Furthermore, the alignment of the different 3D images (the alignment phase) in order to cover the object's surface is also time-consuming. This phase requires significant effort and will affect the final quality (resolution, noise level and accuracy) of the 3D model. During this phase, either device-based (spheres, targets, etc...) or data-based (manual or automatic) initial alignments are used to first generate an approximate model. This is always followed by a more precise Iterative Closest Point (ICP) technique (Soucy et al., 1996). In most cases, the alignment is still a manual process. After the alignment has been successful, the merging phase creates a seamless 3D model that can be edited. Texture images acquired by a separate colour digital camera or a multi-spectral device can be registered (pose calculation) with the 3D model using common points between the 3D model and the 2D image; the texture is then draped around the digital model (El-Hakim and Beraldin, 2007). Here, either a pre-calibration or calibration using 3D points directly on the 3D model are needed to ensure accurate mapping or draping of textures onto a 3D model. Some laser scanners acquire the Red-Green-Blue or RGB information directly and are in perfect registration with the geometry. More details of the 3D model acquisition pipeline can be found in the original paper by (Soucy et al., 1996). Figure 1 outlines the main steps required to create a 3D model from 3D imaging systems. It illustrates the methodology used in constructing coloured 3D triangulated meshes of an object or site. The dotted line indicates the use of other technologies to complement the 3D modelling sequence. For instance, photogrammetry may assist in providing accurate image alignment information within a global alignment framework.

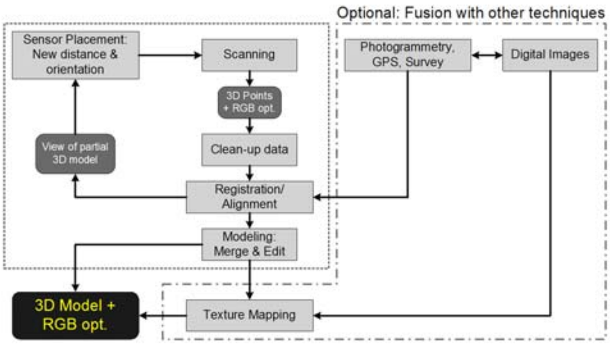

Figure 1. General procedure for 3D modelling; from (El-Hakim and Beraldin 2007).

Many authors (also users of the technology) have worked on methodologies and tools to reduce the time spent on creating a 3D model. In particular, (Andreetto et al., 2004) propose a technique to automate the pre-alignment of 3D images. Their method allows for an automatic pair wise alignment of images as they are acquired in a continuous chain. The set of pre-aligned 3D images is then sent to a global alignment method to finalize the alignment. The other modelling steps can continue as explained above. Such an approach can save time and reduce operator errors.

The approach we propose addresses a situation where many images (2D and 3D) are available, but the 3D modelling has not been started or not been completed (see Introduction). Our approach can be used in a typical pipeline of events as is the case with current commercial software (manual pair wise pre-alignment) or inserted just before an automated method as described by (Andreetto et al. 2004).

3. CONTENT-BASED INDEXING AND RETRIEVAL OF COLOUR AND TEXTURE

Images are difficult to describe (Ki Tae and Young Shik, 2006). They convey a large amount of complex and ambiguous information. The ambiguity is due to the fact that an image is a bidimensional projection of the three-dimensional world and by the fact that the illumination of this world is arbitrary and cannot be controlled. Because of this ambiguity and complexity, it is difficult to segment images and to understand them. For the above-mentioned reasons, we propose a statistical approach in which the overall composition of the image is described in an abstract manner.

We now depict our algorithm. The colour distribution of each images is describes in terms of hue and saturation. This colour space imitates many characteristics of the human visual system. The hue corresponds to our intuition of colour e.g. red, green or blue while saturation corresponds to the colour strength e.g. light red or deep red.

Next, a set of points is sampled from the image. A quasi-random sequence generates the points. In the present implementation, the Sobol sequence has been selected. Each point of this sequence becomes the centre of a structuring element. For each centre position, the pixels inside the corresponding structuring element are extracted and the associated hue and saturation images are calculated. The statistical distribution of the colours within the window is characterized by a bidimensional histogram. The first dimension of this histogram corresponds to the hue or the saturation quantified on a discrete and finite number of channels. The second dimension corresponds to the relative proportion of each channel within the window. This bidimensional histogram is computed and accumulated for each point of the sequence, i.e. the current histogram is the sum of the histograms at the current and at the previous position. From this process, a compact descriptor or index is obtained.

This index provides an abstract description of the composition of the image i.e. of the local distribution of colours throughout the image. This is very important. This index does not represent a global description of the image nor is it based on a particular segmentation scheme. Instead, it characterizes the statistics of colour distribution within a small region that is moved randomly over the image. Consequently, there are no formal relations in between the different regions, which means that the different components of a scene can be combines in various ways while still be identified as the same scene. That is why that algorithm is robust against occlusion, composition, partial view and viewpoint. Nevertheless, this approach provides a good level of discrimination.

As we know, an image is worth a thousand words, which means that it is difficult to describe an image based solely on words. For that reason, our retrieval approach is based on the so-called "query by example" or "query by prototype" paradigm. To this end, we created a search engine that can handle such queries. In order to initiate a query, the user provides an image or prototype to the search engine. This prototype is described or indexed and the later is compared with a metric to a database of pre-calculated indexes, which correspond to the images of the virtual collection. The search engine finds the most similar images with respect to the prototype and displays them to the user. The user then acts as an expert: he chooses the most meaningful image from the results provided by the search engine and reiterates the query process from the chosen image. The process is repeated until convergence is achieved.

4. CONTENT-BASED INDEXING AND RETRIEVAL OF RANGE IMAGES

The indexation of three-dimensional artefacts differs fundamentally from the indexation of images (Ki Tae and Young Shik, 2006, Paquet and Viktor, 2005). If the three-dimensional information has been acquired accurately at a sufficiently high resolution, the three-dimensional geometry constitutes an unambiguous body of information in the sense that there is a one-to-one correspondence between the virtualised geometry and the physical geometry of the artefact. As explained in the previous section, the situation is entirely different for images.

Shape also constitutes a language in its own right. In addition to verbal language, humanity has developed a common shape language. This is particularly evident in fields like art and architecture. For that reason, the "query by prototype" approach is a powerful paradigm for the retrieval of similar artefacts. As far as the overall structural design is involved, the three-dimensional artefact retrieval system is very similar to its image counterpart: the artefacts of the collection are indexed offline and a database of indexes is created. In order to interrogate this database, the query is initiated with a prototype artefact. From the proto-artefact, an index is calculated and compared, with the help of a metric, to the indexes of the collection in order to retrieve the most similar artefacts in terms of three-dimensional shape. As stated before, the user can act as an expert in order to reiterate the process until convergence.

Consequently, the main differentiation between the two systems (image versus 3D) is the index. We now describe our algorithm for three-dimensional artefact description. We assume that each artefact has been modelled with a mesh. This is a non-restrictive representation for virtualised artefact since most acquisition systems generate such a representation. In the present case, a triangular mesh representation is assumed. If the mesh is not triangular, the mesh is tessellated accordingly. Our objective is to define an index that describes an artefact from a three-dimensional shape point of view and that is translation, scale and rotation invariant. The later invariants are essential because the artefact can have an arbitrary location and pose into space.

The algorithm can be described as follows (Paquet and Viktor, 2005). The centre of mass of the artefact is calculated and the coordinates of its vertices are normalised relatively to the position of its centre of mass. Then, the tensor of inertia of the artefact is calculated. This tensor is a 3 x 3 matrix. In order to take into account the tessellation in the computation of these quantities, we do not utilise the vertices per se but the centres of mass of the corresponding triangles; the so-called tri-centres. In all subsequent calculations, the coordinates of each tri-centre are weighted with the area of their corresponding triangle. The later is being normalised by the total area of the artefact, i.e. with the sum of the area of all triangles. In this way, the calculation can be made robust against tessellation, which means that the index is not dependent on the method by which the artefact was virtualised: a "sine qua non" condition for real world applications. In order to achieve rotation invariance, the Eigen vectors of the tensor of inertia are calculated. Once normalised, the unit vectors define a unique reference frame, which is independent on the pose and the scale of the corresponding artefact: the so-called Eigen frame. The unit vectors are identified by their corresponding Eigen values.

The descriptor is based on the concept of a cord. A cord is a vector that originates from the centre of mass of the artefact and that terminates on a given tri-centre. The coordinates of the cords are calculated in the Eigen reference frame in cosine coordinates. The cosine coordinates consist of two cosine directions and a spherical radius. The cosine directions are defined in relation with the two unit vectors associated with the smallest Eigen values i.e. the direction along which the artefact presents the maximum spatial extension. In other words, the cosine directions are the angles between the cords and the unit vectors. The radius of the cords are normalised relatively to the median distance in between the tri-centres and the centre of mass in order to be scale invariant. It should be noticed that the normalisation is not performed relatively to the maximum distance in between the tri-centres and the centre of mass in order to achieve robustness against outliers or extraordinary tri-centres. From that point of view, the median is more efficient that the average. The cords are also weighted in terms of the area of the corresponding triangles; the later being normalised in terms of the total area of the artefact.

The statistical distribution of the cords is described in terms of three histograms. The first histogram describes the distribution of the cosine directions associated to the unit vector associated with the smallest Eigen value. The second histogram denotes the distribution of the cosine directions associated with the unit vector associated with the second smallest Eigen value. The third histogram refers to the distribution of the normalised spherical radius as defined in the previous paragraph. The ensemble of the three histograms constitutes the shape index of the corresponding 3D range image.

5. CONTENT-BASED AIDED REGISTRATION FOR COLOUR AND TEXTURE

The acquisition of a three-dimensional point with a laser scanner is always accompanied with the corresponding acquisition of a colour point. The colour of such a point is related to the wavelength of the laser beam of the scanner.

The most common case is, of course, an infrared laser beam for which the corresponding colour point is monochromatic. Some laser scanning techniques support multiple wavelengths. For instance, the NRC synchronised colour laser scanner (Blais 2004) makes use of a beam constituted of three wavelengths located in the red, green and blue region of the visible spectrum. The resulting laser beam is white which means that it is possible, in addition to the geometry, to acquire the RGB colour of the point.

In addition to this colour per vertex acquisition scheme, it is possible to take digital pictures of the artefact under consideration. Subsequently, the pictures are mapped or draped on the corresponding geometry. In this particular case, the colour is not represented on a per vertex basis, but rather as a set of texture maps with the corresponding texture coordinates. In any case, one has, in addition to the geometry, a colour image associated with each view. In the present section, we propose to take advantage of the chromatic information in order to find range images that should be registered together i.e. range images that are similar, adjacent or complementary. We first describe our method. The later is then applied to the "Madonna and Child", a sculpture made by Pisano (1250-1314) that can be found in the Capella degli Scrovegni in Padova (Italy).

The range data and monochromatic images were acquired in 1997 using a portable 3D imaging system (Beraldin et al. 1998). More than 750 acquisitions were made during a one day session when laptop computers had only 8 MB of RAM. Back then, it was not possible to acquire 3D images and perform on-site 3D alignment. Images were acquired in a systematic way with 50% to 90% overlap between views. Most of the annotations were hand-written in three languages: English, French and Italian. A proprietary 3D file format was used which is no longer available. This format contained very little information on the scanned areas and many distinct files had the same file name. In the present paper, monochromatic images, acquired with a red laser, are used while keeping in mind that the proposed method is not limited to a single colour.

This case study is therefore a perfect example of a multilingual, multinational project. It follows that the underlying methodologies and technologies used 10 years ago, as mentioned above, differ substantially from the current state of the art. Also, the documentation was recorded rather informally and in an "ad hoc" way, making it difficult to accurately decipher the annotation. To address these shortcomings, we use our content-based approach to aid the registration process of related images, as follows.

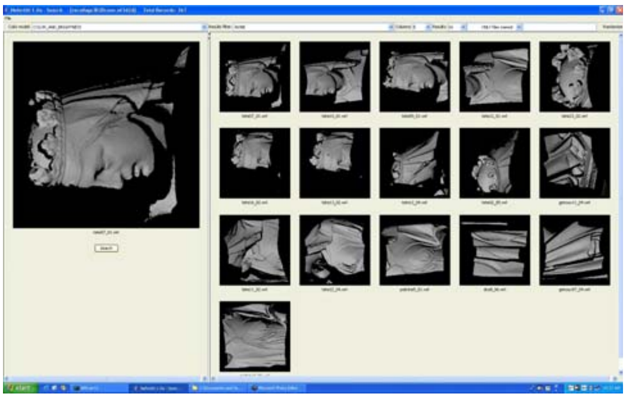

Figure 2. Retrieval of various views of the head from an infrared image.

Our method is based on the algorithm described in section three. In order to utilise this algorithm, we need to generate, for each range image, a corresponding colour image. Such a task is performed as follow. We read the file associated with the range image and we parse the geometrical and colour information. The parsed information is converted into an internal representation.

The geometry and the colour of the internal representation are rendered in three dimensions with our system; with synthetic lighting if required for better visualisation. Then, a colour picture of the rendered view is taken with a virtual camera. Such a picture constitutes the colour image. Once the colour image has been obtained, a descriptor or index that described the chromatic content as well as the composition is generated by the algorithm described in section three. The process is repeated for each view and is entirely automatic. Once all the views have been processed, a database of the descriptors is created. The later may subsequently be searched with our query by example search engine.

Recall that the search engine is based on the query by example paradigm. A picture, which constitutes the example, is chosen by the user. The search engine then utilised the descriptors in order to retrieve the most similar pictures from the database and display *N* results where *N* is determined by the user. For instance, figures 2 to 4 show three queries that were performed for the head and the breast, where *N* is equal to 16. In general, the retrieved pictures correspond to adjacent or complementary views. A mismatch is always possible but, as opposed to the un-aided case, the user is working with a manageable and suitable number of possibilities (typically of the order of ten) instead of having to manage the entire database, which may contain thousands of views.

Figure 2 shows the retrieval of a head. The present example considers a profile of the later in which the crown is visible. The first twelve items (except item ten) are different views of the head that either correspond to or are adjacent to the later. Naturally, it is not possible to retrieve all the views from a single query. However, the other views can be reached with a few iterations if the process is repeated with other examples. For instance, these views can be selected from the outcome of the query that was obtained at the previous iteration.

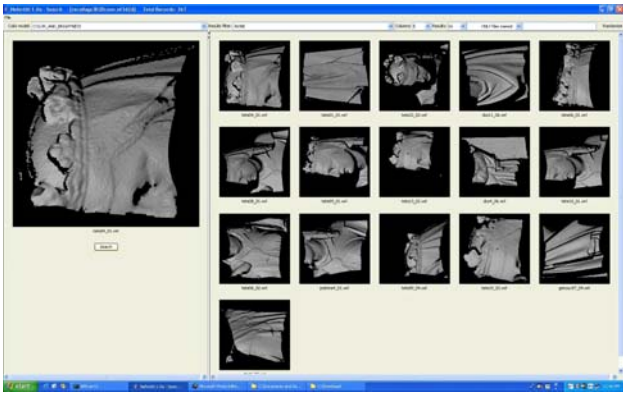

Figure 3. Retrieval of various views of the head from a red image.

Figure 3 shows a query similar to the one effectuated in figure 2. The example also represents another profile of the head in which the veil underlying the crown is more visible. The outcome of the query is constituted of various views of the head and the crown as well as a few outliers that can be easily eliminated by visual inspection. The point of importance here is that, by using a different example, a new subset of the solution space can be reached.

By performing a few iterations, it is possible to find the vast majority, if not the entirety, of the views that are related to the region under consideration.

Finally, we would like to focus our attention on figure 4 which represents once more a profile of the head, a sub-region of the breast as well as a very partial view of the hand holding the child (at the bottom of the image). The search engine has retrieved not only many views of the face but also many views of the breast (pictures 3, 8, 9, 10 and 11) and two views of a hand (picture 11 and 13). In the later case, the user validation is of particular importance since the hand in the query example is not the same that appears in these results. Nevertheless, it is remarkable that the search engine was able to find a hand with an example in which the later is so marginally defined.

These examples, as well as a multitude of others that are not shown in the present paper, indicate that our image-based retrieval technique can efficiently retrieve related or complementary views of the same geometry. We would like to reiterate that the indexing process is entirely automatic and does not require any segmentation or any prior knowledge. All the calculations were performed on 10 year old noisy data without any noise reduction or any other kind of pre-processing. Let us mention that a query is typically resolved in a period of the order of a second (on a Dell Precision M60 laptop). Consequently, a user can make sense of a huge project in a relatively short amount of time, without having to extensively study the original annotation and other documentation, if available. This implies that, for example, a user is able to analyze data that was recorded in a foreign language using a methodology she (or he) is not familiar with. Also, data that were recorded in the past, using outdated technologies, can be processed using the state of the art. Once the adjacent and complementary views have been found, the registration process can proceed as described in section two, again using the methodologies and technologies the user currently prefers and are most familiar with.

Figure 4. Retrieval of various views of the head and the breast from a red image.

6. CONTENT-BASED AIDED REGISTRATION FOR RANGE IMAGES

The framework developed in the previous section can also be successfully applied, in essence, to the registration of range images. The only difference being the descriptor or index which, in the present case, describes the three-dimensional shape of the associated range image. Such an index was described in detail in section four.

As in the previous section, the search engine is based on the query by example paradigm. A 3D shape or range image, which constitutes the example, is selected by the user. The search engine then utilises the 3D descriptors in order to retrieve the most similar 3D shapes from the database and display *N* results where *N* is determined by the user. For instance, figures 5 to 7 show three queries that were performed for the head and the breast, where *N* is equal to 16. In general the retrieved range images correspond to adjacent or complementary views. A mismatch is always possible but, as opposed to the un-aided case, the user is working with a manageable and suitable number of possibilities (typically of the order of ten) instead of having to manage the entire database which may contain thousands of views.

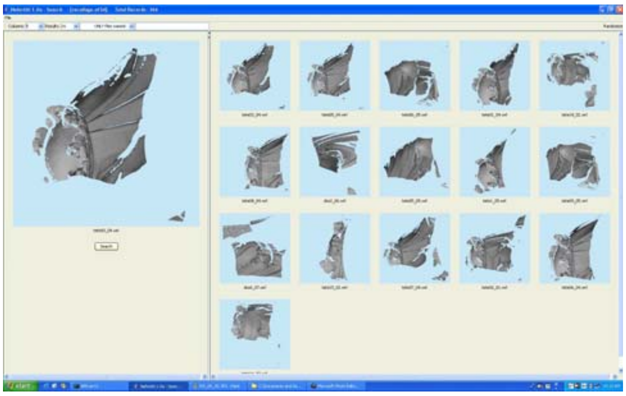

Figure 5. Retrieval of various views of the crown from a range image.

Figure 5 shows the retrieval of the back section of the head in which the crown and the veil are pre-eminent. The outcome of the query is made of various views of the crown and the veil. There are no outliers. Such a good result can be explained by the fact that 3D shape retrieval, as opposed to 2D image retrieval, is not ambiguous in the sense that a 3D shape is the result of a measurement of the real physical geometry, while a 2D image is merely a projection of the later on a surface.

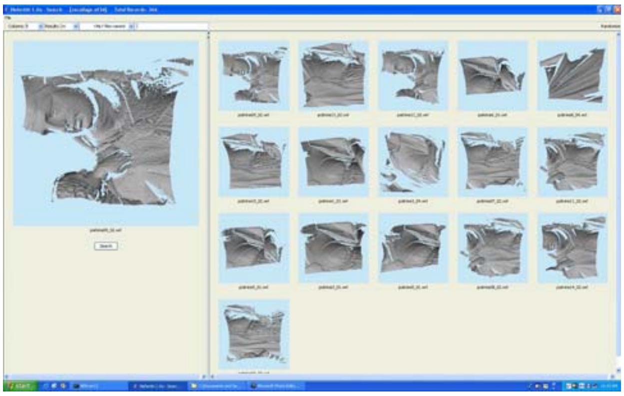

Figure 6. Retrieval of various views of the head and the breast from a noisy range image.

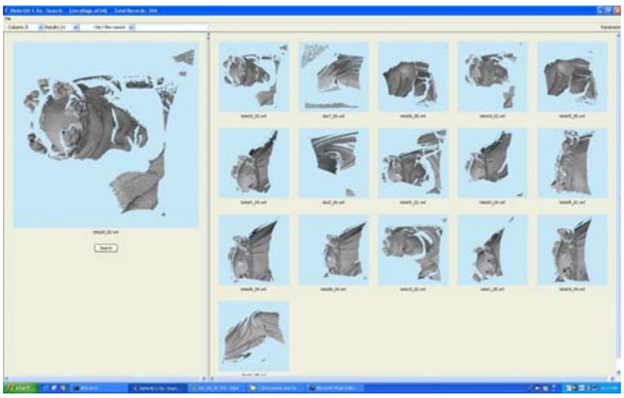

Figure 7. Retrieval of various views of the crown and the face from a noisy range image.

Figure 6 shows the retrieval of the breast and the face. The outcome of the query comprises range images that are related to the breast. There are no outliers for the reasons described earlier. Finally, figure 7 shows the retrieval of the face and the crown. As in the previous examples, various adjacent or related range images are extracted from the database.

The illustrative examples show that our 3D shape retrieval technique is able to efficiently retrieve related, or complementary, views of the same geometry. As in the previous section, the indexing process is entirely automatic and does not require any segmentation or any prior knowledge. All the calculations were performed on noisy data without any noise reduction or any other kind of pre-processing. This, again, makes it highly suitable for users who were not necessarily involved in the acquisition project.

7. CONTENT-BASED AIDED PHOTOGRAMMETRIC PROJECT

In this section, we address the search and retrieval of similar and related digital images from a database of 100.000 views of the "Conversion of St Augustine" and shows how two different, complimentary query representations (with and without spherical distortion) of the same painting may be utilized by our method. Here, our aim is twofold. Firstly, we want to study the case in which the fresco would be painted on a spherical cupola. Secondly, we are considering the situation where two teams have completed their surveys at two points in time, and the representations of one survey is distorted due to aging or bad preservation.

Figure 8 shows a high-resolution digital photograph (4.000 x 4.000) of the "Conversion of St-Augustine" which was painted by Fra Giovanni da Fiesole (1387-1455). St-Augustine is shown crying after his conversion in a garden.

Figure 8. "Conversion of St Augustine" painted by Angelico Fra Giovanni da Fiesole (1387-1455).

Firstly, we consider the painting without any spherical distortion. From this painting; we generated 100.000 fragments as follows. A high resolution version of the painting (16.000 x 16.000) was first mounted on a virtual frame and pictures of sub-images were taken from various orientations and distances with a virtual camera. The azimuthal angle can take any value in between 0^o and 360^o and the altitude angle any value in between 0^o and 30^o. The distances were chosen such that the area of a given fragment would correspond to a fraction of the area of the original; in the present in between 1 and 5 %.

The appearance and the composition of the fragments were indexed with the method described in section 3. The first set of results is presented in figures 9 to 14, where *N* was set to 30. Figure 9 shows the retrieval of one of the donator. A recall of 100% is obtained with a precision of 100%. This means that all the related views are retrieved and that there are no outliers. This is remarkable considering that the database contains 100.000 images. This is consistent for all the results presented here. As it can be seen, the donator is retrieved irrespectively of the viewing angle.

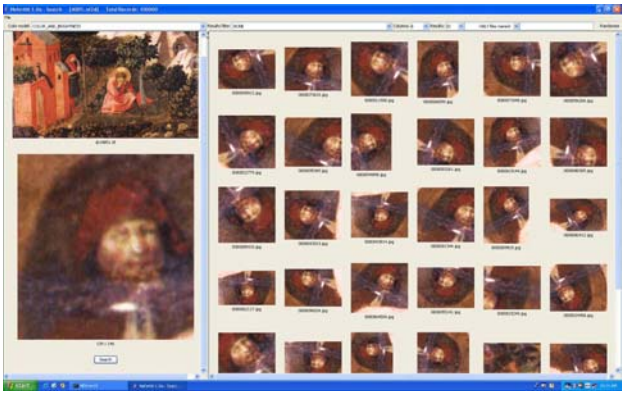

Figure 9. Retrieval of various views of the donator.

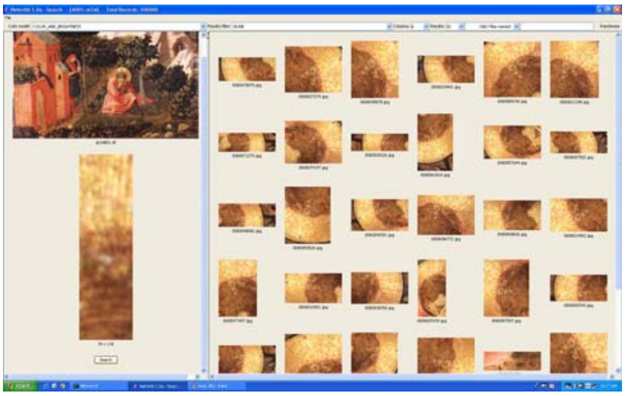

Figure 10. Retrieval of various views of a sub-image of Augustine's ring.

Figure 10 shows the retrieval of a sub-image of St-Augustine's ring. Despite the fact that the resolution of the reference sub-image is very low, all the rings were retrieved.

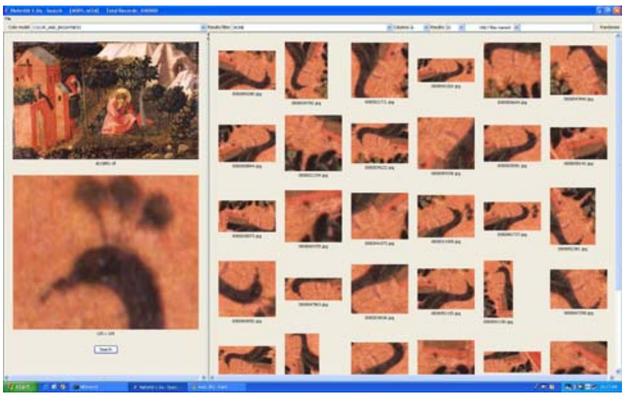

Figure 11. Retrieval of various views of the peacock head.

Figure 11 shows the retrieval of the head of the peacock. This example illustrates very well the robustness of the method against the viewing angle. As for all the other examples presented here, a recall of 100% is obtained with a precision of 100 % for the first 30 images retrieved.

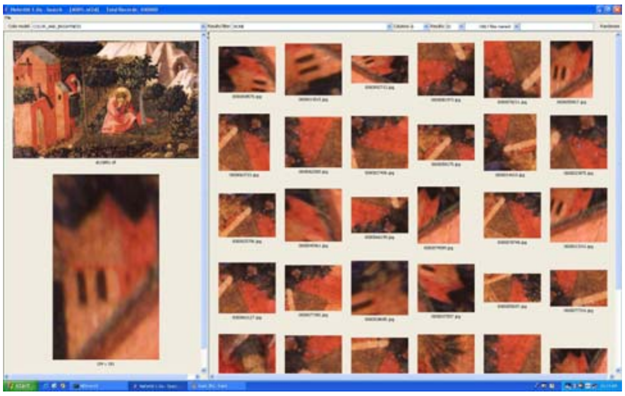

Figure 12. Retrieval of various views of building in the background (top-left corner).

Figure 12 shows the retrieval of a building situated in the background as located in the top-left corner of figure 8. The system again managed to retrieve all the views despite the fact that the query image has a very low resolution. One should notice that the background building has not been confused with the main building, despite a very similar appearance.

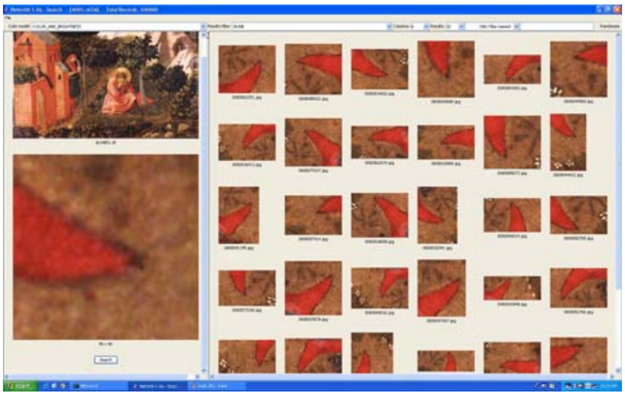

Figure 13. Retrieval of various views of a shoe.

Finally, figure 13 shows the retrieval of the end of a shoe of the other donator of the painting. All the views of the shoes were retrieved without any confusion with the background building despite the fact that some views are relatively similar, as can be seen when one compares figures 12 and 13.

Next, we consider the case where the reference image contains spherical distortion e.g. a fresco painted on a spherical cupola. Distortions also appear as a result of time or bad preservation which means that the original "flat" painting can progressively take a distorted shape. It follows that this should be recognizable by the retrieval system. This means that the system should be robust against distortion in order to be viable and useful in long-term projects. Figure 14 shows the Conversion with a spherical distortion. Figures 15-17 show that is possible to find similar or related images not only if the reference image is flat e.g. a fresco painted on a flat wall, but also if it is curved e.g. a fresco painted on a cupola. Consequently, in the following examples, the query images were extracted from the distorted painting. The same fragment database (100.000) was utilised, since the curvature of the fragments is negligible due to their reduced dimensions and their corresponding physical small size on the painting.

The system was unaffected by the distortion and the usual recall and precision were obtained even in the zones of very high distortion. Figure 15 shows the retrieval of the end of the tail of the peacock. The later is in a zone of high distortion. The results are quite remarkable in the sense that the tail was not confused with the window of the main building despite of their very similar appearance.

Figure 14. "Conversion of St Augustine" painted by Angelico Fra Giovanni da Fiesole (1387-1455) with simulated spherical distortion.

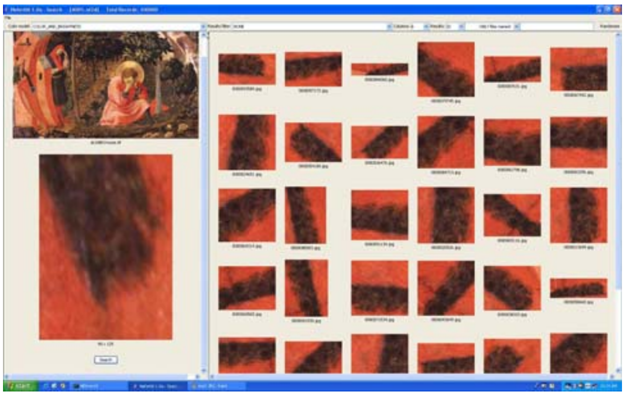

Figure 15. Retrieval of various views of the tale of the peacock from the distorted image.

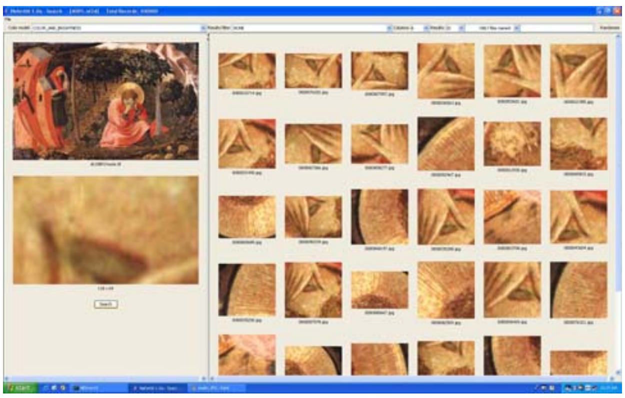

Figure 16. Retrieval of various views of Augustine's eye from the distorted image

Figure 16 shows the retrieval of the eye and two fingers of St-Augustine. The resolution of the reference image is very low but the related fragments were properly identifies with the usual precision and recall rate.

Finally, a part of the head of the peacock was selected as a query, as shown in figure 17. All parts of the head were retrieved even though the head is in a zone of high distortion. This result should be compare with figure 11 which presents a similar case but with no distortion. Part of the robustness of the method comes from the fact that it is pixel-based and not contour or segmentation-based. All the queries performed in this paper were completed within a second.

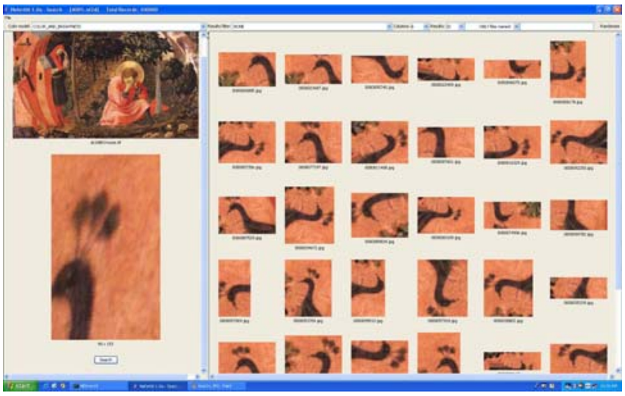

Figure 17. Retrieval of various views of the peacock head from the distorted image. The results are comparable to the non-distorted case illustrated in figure 11.

The examples show that from a (or many) picture(s) of an original, it is possible to search and retrieve very efficiently similar or related images from very large databases (in the present case 100.000). It follows that such an approach could be used, besides, to manage large photogrammetric projects. One can find, for instance, a picture from another survey that can fill a hole in the model or a set of pictures to cover a particular detail of relevance that is currently missing. Importantly, the presented approach allows one to combine seamlessly recent surveys together with older one and historical images acquired using diverse underlying methodologies and multilingual annotations. It is possible to process or index them either separately or together and to then combine the databases for retrieval purpose. In this case, it follows that the robustness against lighting conditions and colour calibration may be problematic during data integration. However, as was shown in (Paquet et al. 2006) with a database of 162.000 images, our method has proven itself to be robust under such conditions. This issue will be further explored in future work.

8. CONCLUSIONS

We have presented a content-based retrieval aided registration method for textures and range images, for use by multilingual and multinational teams. This method allows to automatically describe the appearance and the geometry of each view and to search for the most similar and related views for registration within the acquisition database. The method does not need any human intervention, except for validation, and is language and methodology independent making a suitable candidate for large scale multinational projects that spans many years. Importantly, it provides users with the ability to freshly access surveys from the past using the state of the art. Furthermore, the method handles noisy data well and does not require any data pre-processing.

9. AKNOWLEDGMENTS

The authors would like to thank Shawn Peters and Luc Cournoyer for their precious collaboration.

10. REFERENCES

1. Andreetto, M., Brusco, N. and Cortelazzo, G. M., 2004. "Automatic 3D modeling of textured cultural heritage objects", IEEE Transactions on Image Processing, (13) 3, 354-369.

2. Beraldin, J.-A., Blais, F., Rioux, M., and Cournoyer, L., 1998. "Portable digital 3D imaging system for remote sites," Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '98) Conference, May 31-June 3, Monterey, CA, USA.

3. Blais, F., 2004. "A review of 20 years of range sensor development", Journal of Elect. Imaging, 13 (1), 231-243.

4. El-Hakim, S. F., and Beraldin, J.-A., 2007. "Sensor Integration and Visualisation, Chap.10 in "Applications of 3D Measurements from Images" edited by J. Fryer, A. Mitchell and J. Chandler, Whittles Publishing, UK.

5. Iyer, N., Jayanti, S., Kuiyang, L., Kalyanaramanand, Y. and Ramani, K., 2005. "Three-dimensional shape searching: state-of-the-art review and future trends", Computer-Aided Design, (37) 5, 509-530.

6. Ki Tae, P. and Young Shik, M, 2006. "Object-based image retrieval using dominant color pairs between adjacent regions", Computational Science and its Applications – ICCSA 3984, 404-411.

7. Paquet, E. and Viktor, H. L., 2005. "On the application of 3-D technologies to the framework of cultural heritage", Invited Paper, Annales des télécommunications 60, 1311-1325.

8. Paquet, E. and Beraldin, J.-A., 2007. "Content-based Retrieval Aided Registration for 2-D Images and Range Images", SPIE Electronic Imaging 2007, Proc. SPIE Vol. 649, 28 January – 1 February, San Jose, USA, 64910F-1-64910F-10.

9. Paquet, E., Lahanier, C., Pitzalis, D., Peters, S., Aitken, G. and Viktor, H. L., 2006. "Content-based Indexing and Retrieval of Cultural Heritage Data: an Integrated Approach to Documentation with Application to the EROS Database", The 7th International Symposium on Virtual Reality, Archeology and Cultural Heritage – VAST 2006, EPOCH Publication, The e-volution of Information Communication Technology in Cultural Heritage, Where Hi-Tech Touches the Past: Risks and Challenges for the 21st Century, Project papers from the joint event CIPA/VAST/EG /EuroMed 2006, 30 October - 4 November, Nicosia, Cyprus, 270-275.

10. Soucy, M., Godin, G., Baribeau, R., Blais, F. and Rioux, M., 1996. "Sensors and algorithms for the construction of digital 3-D colour models of real objects", Proc. Intern. Conf. on Image Processing, vol. II, Sept. 16-19, Lausanne, Switzerland, 409-412.