ACQUISITION AND REPRODUCTION OF SURFACE APPEARANCE IN ARCHITECTURAL ORTHOIMAGES

A. Martos[a],[*] R. Cachero[a]

[a]Dogram, Parque Tecnológico de Asturias, 33428 Llanera, Spain - (martos, rocio)@dogram.es

Commission V, WG V/4

KEY WORDS: Orthoimage, Multi-View Stereo, Photometric Stereo, Surface Appearance, BRDF Capture, Physically Based Render

ABSTRACT:

Software tools for photogrametric and multi-view stereo reconstruction are nowadays of generalized use in the digitization of architectural cultural heritage. Together with laser scanners, these are well established methods to digitize the three-dimensional geometric properties of real objects. However, the acquired photographic colour mapping of the resulting point clouds or the textured mesh cannot differentiate the proper surface appearance from the influence of the particular illumination present at the moment of the digitization. Acquisition of the actual *surface appearance*, separated from the existing illumination, is still a challenge for any kind of cultural heritage item, but very specially for architectural elements. Methods based on systematic sampling with commuting light patterns in a laboratory set-up are not suitable. Immovable and outdoor items are normally limited to the existing and uncontrolled natural illumination.

This paper demonstrates a practical methodology for appearance acquisition, previously introduced in (Martos and Ruiz, 2013), applied here specifically for the production of re-illuminable architectural orthoimages. It is suitable for outdoor environments, where the illumination is variable and uncontrolled. In fact, naturally occurring changes in light among different images along the day are actually desired and exploited, producing an enhanced multi-layer dynamic texture that is not limited to a frozen RGB colour map. These layers contain valuable complementary information about the depth of the geometry, surface normal fine details and other illumination-dependent parameters, such as direct and indirect light and projected self-shadows, allowing an enhanced and re-illuminable ortoimage representation.

1. INTRODUCTION

1.1 State of the art

Nowadays the use of diverse software tools for photogrammetric and image-based dense 3D reconstruction is widely generalized in cultural heritage digitization. There are quite effective methods to automatically acquire and reproduce the shape and colour texture of real cultural heritage objects. The so-called Multi-View Stereo (MVS) or Image Based Modelling (IBM) methods depart only from multiple images, normally taken under existing (passive) illumination conditions. When properly applied they can typically reconstruct the original geometry, reaching a considerable level of detail and accuracy, competing with other (active) methods such as laser scanning. Other active scanning methods such as structured light scanning are of limited use when outdoors or applied for large architectural elements, mainly due to practical limitations on the intensity of the light.

In any of these cases the colour of the surfaces is intended to be reproduced by means either of coloured point clouds or static texture maps. Both are generated by projection of the images. In some cases the colour is quite directly extracted from the photographs by projecting the pixels of each image onto the 3D model surface polygonal mesh, creating high resolution blended colour textures. When using a laser scanner, complementary photographs are typically taken on the spot as well, using a separate or an integrated camera. This provides either a coloured point cloud by back-projecting the points onto the images. Both coloured models can be used to straightforwardly produce conventional orthoimages by simply rendering a projection of the 3D model onto an orthogonal plane of choice.

[*]Corresponding author

One obvious limitation of these methods resides in the fact that even when the objects are accurately photo-textured, and regardless of the method used to map the colours, the reproduced surface texture of the orthoimage will be a static snapshot frozen onto the surface of the object, resembling only the existing conditions at the moment of the acquisition. When the images are taken in a cloudy day the variation of illumination among different images is normally small, favouring thus the matching and reconstruction. However the resulting representation of colour texture is perceived as flat, lacking much of the potential contrast and detail due to the limited shading (Fig. 1).

Figure 1: Portada de la Oliva, Asturias, Spain. Orhtoimages of the same object generated from photographs taken either in cloudy conditions (left) or under direct sunlight (right).

Whereas under direct sunlight the surface shading and self-projected shadows will be much vivid and informative. Sharper shading may significantly help in the intuitive interpretation of the image, by giving better appreciation of volumes and fine relief details. Ordinarily all the images in each set should be taken under the

same fixed illumination conditions to reduce mismatches caused by the moveable shadows due to changes in sun position, and to avoid the appearance of artificial seams in the blending of multiple images. This requires the selection of the proper time of the day and also completing the field work in a relatively short period of time, which might not be a practical approach for large and complex elements.

Even when all the images were taken with similar shading, the final resulting appearance will not be able separate the proper surface properties from the fixed effects of shading caused by the existing illumination. Hard shadows will be frozen into the model texture, sometimes even hiding relevant details. In some cases faces oriented north (in the northern hemisphere) may never receive enough sunlight at the appropriate angle.

It is not always feasible to avoid changes in natural light when doing outdoor projects. Nevertheless the intended goal in appearance acquisition is actually to characterize how the surface appears to reacts to changes in illumination, turning these variations in the source images into something indeed necessary to complete the documentation. In a similar way, previous approaches to solve appearance acquisition in the laboratory do actually need to produce variable illumination patterns to characterize the Bidirectional Reflectance Distribution Function (BRDF).

1.2 State of the art in appearance acquisition

The acquisition and realistic reproduction of the surface appearance properties of cultural heritage elements is still an open topic addressed only recently. Several previous efforts have been made in order to model and capture the properties of the surfaces, producing digital models that can be observed and illuminated from any angle, and to try to achieve a fully dynamic photo-realistic quality. Some of these approaches are specific for the capture of bi-directional reflectance information of cultural heritage homogeneous materials. The MERL BRDF database includes samples of several materials recorded by the Rochester Institute of Technology in the USA while only as homogeneous and isotropic samples (Chen et al., 2011).

To acquire a non-homogeneous texture mapping of the surface properties, most of the approaches proposed so far require active and variable illumination set-ups, such as domes (Klein and Schwartz, 2010), (Schwartz and Klein, 2012) or goniometer style arms (Haindl and Filip, 2013) and/or rotating tables. These typically require a very large amount of photographs, taken from different known camera positions and commuting multiple point light sources. While these set-ups have been already successfully into acquiring appearance for small objects that can fit inside the working volume, they are still highly impractical or impossible to use for medium or large size objects such as sculptures or architectural items. These devices cannot be used out of laboratory conditions, nor with unmovable objects or for outdoors acquisition. Due to the size of the objects controlling an active illumination completely is no longer a viable option, severely limiting thus the possibilities of these acquisition processes.

Additionally these approaches aim to very exhaustively record all the information required to completely reproduce the very multi-dimensional sampling space of the (BRDF) texture. This a very systematic but not completely practical approach, requiring very exhaustive field data and heavy data processing, together with accurate prior knowledge of geometric and radiometric calibration, as well as a very controlled and closed light environment.

Other approaches successfully combine dense geometry with smart photographic mapping from multiple images, overcoming thus

some of these limitations (Callieri et al., 2008). The resulting visualization is often considered photo-realistic and has enough freedom of movement for the observation point, as long there are no relative movements between the item and the light. However the actual limitation to achieve a completely realistic dynamic visualization lies in the fact that, at best, all the texture/color information is still frozen onto the object surface, not reacting to any potential changes in the color or position of lights or the environment.

Another interesting approach for the reproduction of reflectance properties of cultural heritage objects is Polynomial Texture Mapping (PTM), a computational photographic method that captures a subject's surface shape and color and enables the interactive re-lighting of the subject from any direction (Malzbender et al., 2001). Each RTI resembles a single, two-dimensional (2D) photographic image, but unlike a typical photograph, it can be re-lighted using a "virtual" light. This changing interplay of light and shadow in the image discloses fine details of the subject's 3D surface form. Digitization procedures are flexible and allow the digitalization of larger artworks, but does not capture the 3D model of the object. Instead it represents a texture image which, unlike a typical photograph, can be re-lighted using a virtual light source. All images must be taken from the same position and since the method does not produce true reconstruction it does not allow for an arbitrary choice of the point of view for orthoimage generation.

The approximate capture of the combined illumination and reflectance model was recently addressed with fixed, but unknown light environmental maps (Palma et al., 2012), using a Phong model with clustering of materials. This work shows that effective field acquisition of some properties of the illumination map and surface appearance properties is feasible even with partially specular objects, and that very simple reflectance models can clearly enhance the appearance perception of the rendered model.

The approach proposed in (Tunwattanapong et al., 2013) approximates the illumination conditions using continuous spherical harmonics, using a complete 3D geometrical model and a set of combined images of the object . This method allows the visualization of a variety of objects difficult to acquire with other techniques.

We focused in developing methods able to simultaneously determine the three-dimensional shape of the object and to map the surface appearance and environment illumination simultaneously as a lightmap, determined indirectly, instead of a set of well known point sources. This would allow to work with hand-held digital photographs taken under variable but flexible illumination (including natural light) and would define an approach to appearance capture as practical and flexible as standard MVS, which can be used successfully in a variety of situations and environments, very specifically for architectural and outdoor objects.

1.3 Field and outdoor acquisition

Working outdoors with large architectural elements or sculptures means that illumination can rarely be controlled much, if at all. Even modelling the light-map is not a trivial task. While the size, position and movement of the sun is absolutely predictable given accurate geo-location and time, this is not the case for the indirect light coming from partly overcast sky, ground and nearby buildings. This indirect illumination will still play a very relevant role in the capture process since it is a significant component of the actual light map (see Fig. 3. To illustrate the adaptation of the appearance capture technique described in (Martos and Ruiz,

2013) to outdoors illumination, we will show following examples, taking advantage of naturally occurring variations along the day.

In practice it is often possible to find more than sufficient variation in the illumination pattern along in a single day. With some proper planning and depending on the weather, it is even possible to perform the acquisition taken the pictures under several significantly different light-maps, in just a few minutes.

Figure 2: Ortoimage generated from the virtual model with appearance capture simulating completely overcast sky. Shadows are very diffuse and there is limited volumetric appreciation.
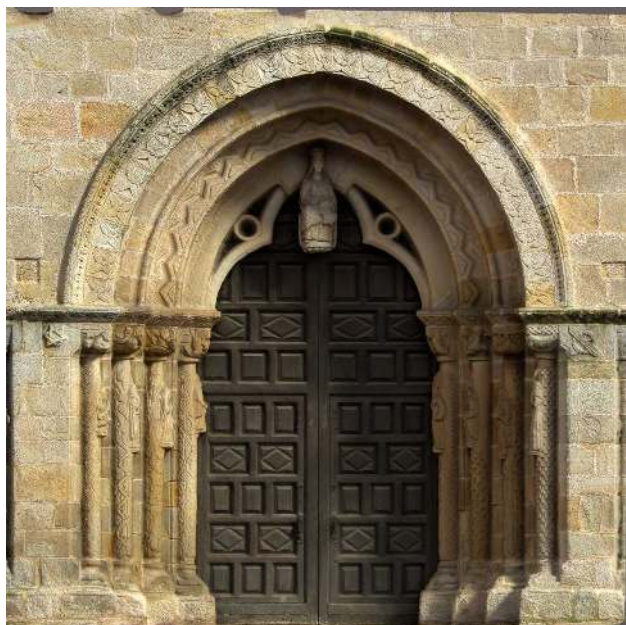
Figure 3: Ortoimage generated from the virtual model with appearance capture simulating direct sunlight with partly overcast sky. Shadows are slightly less sharp than in the original images.

The door of Santa Maria de la Oliva in Asturias, Spain, was photographed at a few different times in the day: at morning, at midday and then again at dusk. Partly cloudy conditions on a windy day in an urban environment provided a large variety of different illumination changes. The most useful set however was obtained in the scope one hour at dusk, when there were rapid changes in sunlight were casting onto nearby structures. Overcast photographs where useful to initialize the texture maps while finer details in texture where obtained from sunny and partly overcast sets of photographs, for which Photometric Stereo methods worked the better.

In any case the selected sets of images should represent a reasonable variation in light. When images are taken under variable illumination conditions, if they are properly grouped and sorted this will not affect matching significantly since differences will not be very significant. However the resulting composition of a single orthoimage from multiple images would exhibit discontinuities in different patches, or result at best in a blended image.

Photographs taken when completely overcast cannot provide much fine texture detail information, but still worked robustly for usual IBM tasks such as feature matching, bundle adjustment, camera calibration and surface reconstruction (Fig. 2). While the resulting mesh was not highly detailed, the additional details are provided by photometric stereo methods with the more direct sunlight images.

Direct sunlight (Fig. 3) also caused saturation in some areas due to the limited dynamic range of the photographs (12 bits). To minimize these problems the photographs where taken with bracketing methods to enhance dynamic range. The light-map estimation has to be calculated in this case with increased resolution (256x256 pixels) and dynamic range (32 bit).

2. METHODOLOGY

In this paper we present new methods for the acquisition and reproduction of surface appearance properties of immovable and outdoor objects, based only on naturally changing illumination conditions. For this acquisition method the sky and surrounding environment will be approximated as a variable and initially unknown light-map, for which changing daily illumination is expected to cause enough variability in the illumination pattern to reproduce the apparent response of the surfaces to the changing illumination.

To illustrate this method we will show how it is was applied to the generation of reilluminable orthoimages of the *Portada de la Oliva* and the door of *Peñalba de Santiago* both architectural outdoor elements with a significant cultural heritage value and interest in detailed representations.

The whole process is achieved departing only from photographs. Several snapshots of the doors where taken using a calibrated digital camera with some particular settings (fixed focus and focal length, manual exposure, low ISO setting, etc...). A basic three-dimensional dense geometry reconstruction is performed by means of dense stereo MVS and SfM methods. This provides dense depth information (Fig. 4) to an scale in the millimetre range which is mapped as a UV texture map (Fig. 5).

A few more photographs where taken within a few hours intervals with varying environmental lighting conditions, including variable sunlight and overcast levels. Using this information a rough illumination model is built and some properties of the BSDF (Bidirectional Scattering Distribution Function) of the surfaces are deduced from the different lighting settings. This allows to produce an specially built bi-directional "pixel shader" function
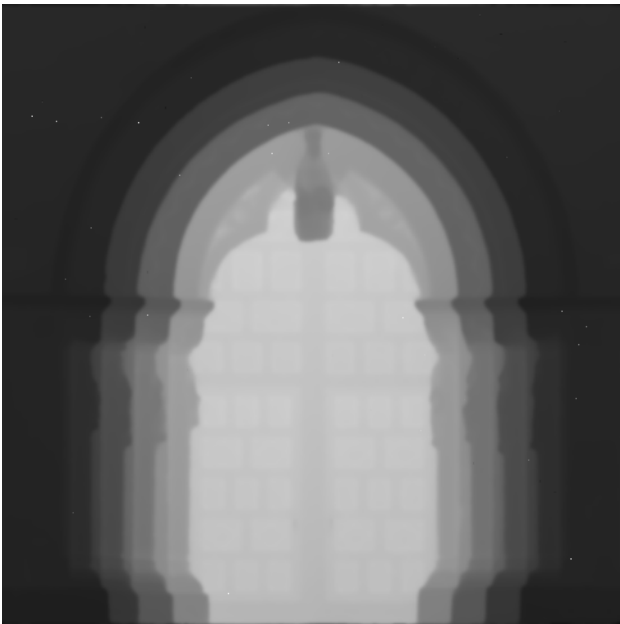
Figure 4: Depth map of the densely reconstructed model. Each pixel represents a different depth w.r.t the chosen orthogonal view plane.

Figure 5: Colour coded UV map, to arbitrarily reassign 3D model coordinates to the 2D texture coordinates of the multiple maps.

to reproduce any new lighting condition including arbitrary new angles for sun incidence.

Additionally fine details in the surface relief are revealed by a normal map (Fig.6) produced by a simplified Photometric Stereo reconstruction method. Direct and indirect shadows, as well as Indirect light scattering are also modelled to properly estimate all illumination contributions in acquisition and to provide a higher degree of realism in the final render. This will provide unified and convenient information about illumination conditions to construct complex ortho-image documents, to enhance depth perception and to provide the volumetric information in the representation that is often missing from conventional planar orthoimages.

We used Patch-based Multi-view Stereo [PMVS] (Furukawa and Ponce, 2010) to reconstruct a dense point cloud with normal mapping, from the set of images and previously calibrated camera internal parameters. The output, a large set of oriented point, is filtered and interpolated with a subsequent matching and hole covering pass, averting the creation of a polygonal (or a mesh) model. This produces both the 3D coordinates and the surface normals estimated at each oriented point which is sufficient for displacement map rendering. However the normal map obtained by this method was shown insufficient for the purposes of detailed surface appearance acquisition. Instead

A natural improvement was to take an initial step where a Lambertian, matte or pure diffusive reflective model is assumed for the surfaces. In this case the apparent color does not depend on the observation angle but only on the incidence of light. It is possible then to determine the light-map by an iterative optimization process based on Photometric Stereo (PS) methods (Basri et al., 2007), used to recover the reflectance properties simultaneously to the refinement of the light-map. Lighting map calculated includes arbitrary combination of diffuse, point and extended sources. General lighting conditions are represented using low order spherical harmonics discretized in the light-map, using simple optimization in a low-dimensional space. This approach is limited when the surfaces are significantly glossy of specular. While this is not the case in this case study, however
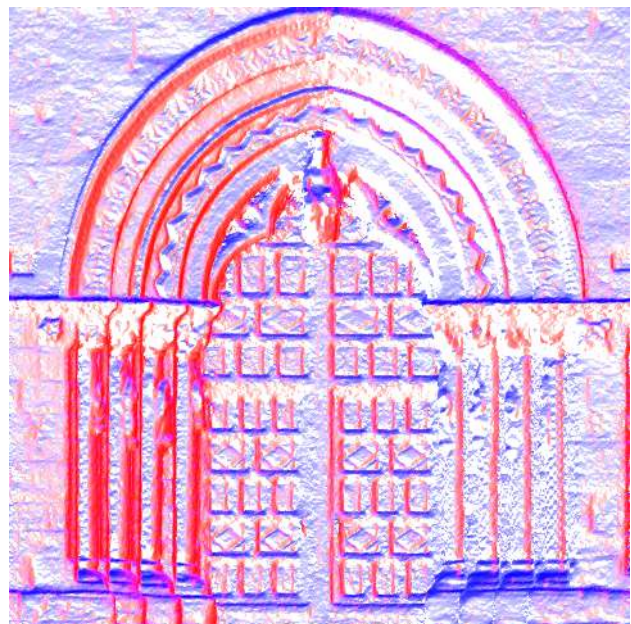

Figure 6: Normal map of the reconstructed model obtained by Photogrammetric triangulation. It was later refined at render time by small corrections acquired by Photometric Stereo methods. Red/blue colours code horizontal/vertical normal components.

this is considered in further refinement stages and the methodology has been tested as well with significantly specular reflections.

These case-studies will show the feasibility of working with field calibration and initial rough approximations for the camera model and light-maps, addressing thus the flexibility required for practical field documentation in museum environments or outdoors. The potential for diffusion will be shown with the use of open source software tools for enhanced visualization. The presented capture methods are integrated with the specific adaptation of open-source GPU-based render engines to produce two flavours of 3D inspection/visualization tools with proper relighting capa-

bilities, able to reveal very subtle details: An quasi-real time realistic engine (Blender Cycles), which is also the basis for the capture process and is focused on realistic reproduction, and a real-time version based on customized *pixel shaders*, for the real-time visualization of lightweight models on web browsers and other interactive applications.

This is achieved not only by reconstructing the shape and projecting colour texture maps from photographs, but also modelling and mapping the apparent optical response of the surfaces to light changes, while also determining the variable distribution of environmental illumination of the original scene. This novel approach integrates Physically Based Render (PBR) concepts in a processing loop that combines capture and visualization. Using the information contained in different photographs, where the appearance of the object surface changes with environmental light variations, we show that it is possible to enhance the information contained in the usual colour texture maps with additional layers. This enables the reproduction of finer details of surface normals and relief, as well as effective approximations of the Bi-directional Reflectance Distribution Function (BRDF). The appearance of the surfaces can then be reproduced with a dedicated render engine providing unusual levels of detail and realism due to enriched multi-layer texture maps and custom shading functions.

The development of these digitalization methods was inspired by several concepts borrowed from relatively recent, but already well developed, 3D rendering methods: The separate capture and modelling of object geometry, reflectance textures and environmental light. The optimizations used by the realistic rendering algorithms, are not just convenient ideas to increase realism under highly constrained render times. Common render pipelines and capture algorithms share fundamental analogies, closing the cycle in order to realistically reproduce digitalized models.

In contrast, we developed a flexible methodology to produce effective surface appearance approximations with a single moving camera and uncalibrated extent lightmaps instead of discrete point sources. This is strongly based on the same principles of free moving camera/station determined by Bundle Adjustment, which is well known for Photogrammetric or Structure from Motion (SfM) methods. The idea is extended here to capture not only the object geometry by triangulation of rays, but similarly, to determine the position of individual light sources which can be approximated with similar principles by uncalibrated stereo-photometric methods. The second key concept exploited is the additivity of light in the image formation process: Light coming from multiple light sources, including indirect or secondary surface reflections in this consideration, is added in the final image, being thus linearly separable. It is thus possible to produce multiple layers of texture and environmental map. This process can be iteratively refined and solved until convergence of the difference between the original and rendered images.

2.1 Observations and limitations

This methodology is characterized by the practical combination of some differential key concepts and methods:

1. The common use of Physically Based Render *shaders* both for surface appearance acquisition and render.

2. Separation of the relevant surface properties in multiple 'render-oriented' layers such as depth, normal, direct and indirect shading, calculated with different methods.

3. The use of an extent light-map, determined indirectly from the images, instead of known point sources.
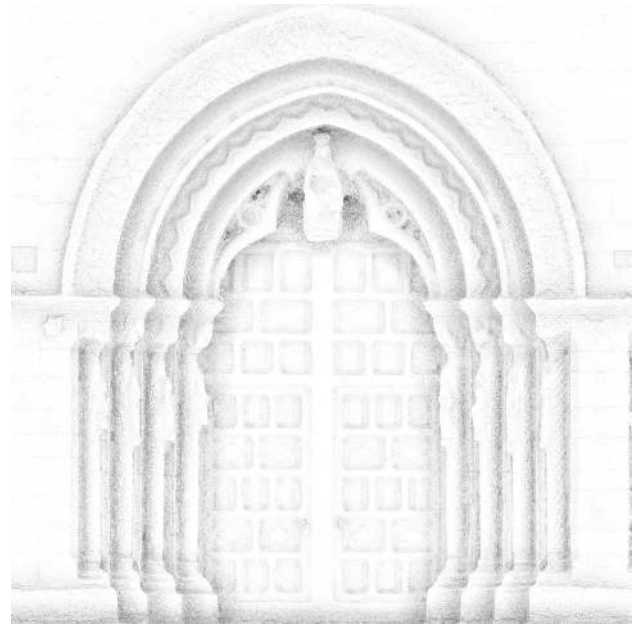

Figure 7: Map with the ambient occlusion shading produced by the diffuse component of the light-map.
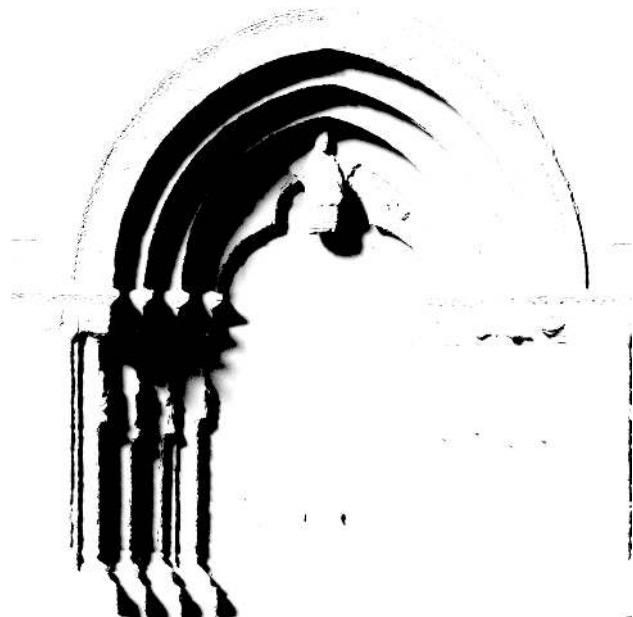

Figure 8: Projected shadows layer, showing areas that will not receive direct light from the sun if set in an arbitrarily given direction. Notice diffuse borders due to sun angular extension.

4. Normal maps calculated by Stereo Photometric methods, more detailed than those calculated by triangulation.

The involved techniques are individually known already and applied separately in cultural heritage digitization and visualization, although not often together. There are actually existing open-source or well documented algorithms for each stage that can be applied to extend the usual MVS pipeline work, by the integration also of PBR and PS methods.

Physically Based Render methods are not only important to provide a more realistic look to the appearance characterization. In-
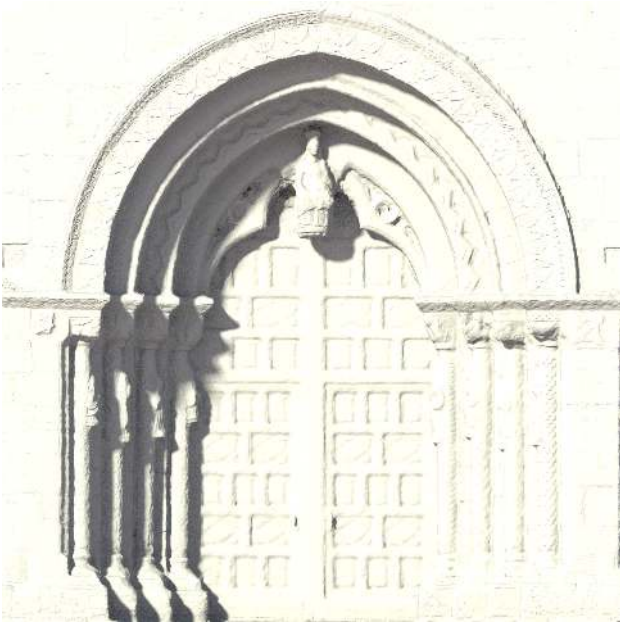
Figure 9: Composition of all the estimated illumination components of one of the images, diffuse, direct and shadow.

Figure 11: Final render of the fully composed orthophoto with a synthetic slanted illumination that highlights the fine relief details of the texture. Enhanced detail is achieved thanks to the photometric stereo method.
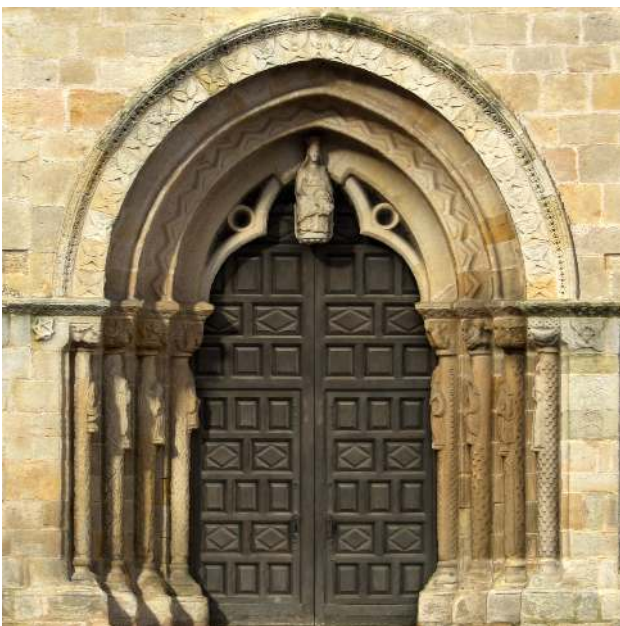

Figure 10: Final orthoimage, made by composition of layers with alternative position for the sun w.r.t Fig.3.
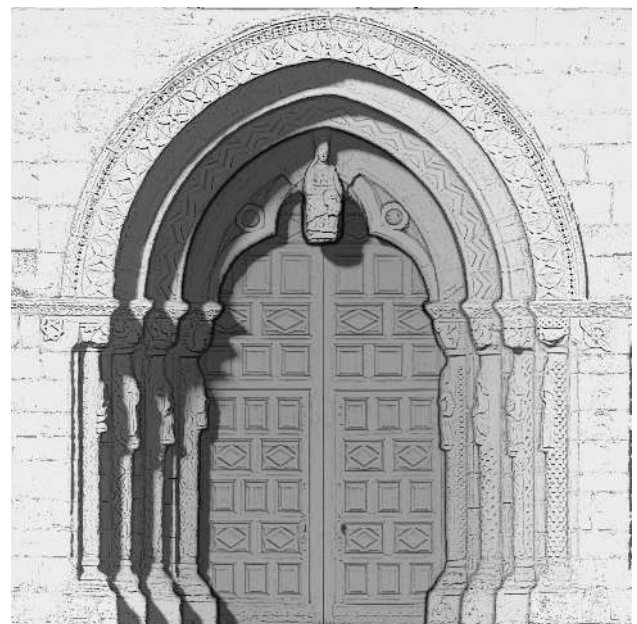

Figure 12: Synthetic drawing created from the processing and composition of multiple layers, removing color information and adding edge and shading, with the purpose showing only the details in shape and fine relief.

direct illumination is indeed a relevant factor in acquisition which cannot be neglected to achieve good results, as shown in figure (Fig.13). Here some surfaces in the object cast light onto nearby surfaces, especially in salient features. This is specially strong when original images have been taken under direct sunlight, producing also hard shadows.

It has been noted as well that concentrated light sources such as the LEDs used in other approaches for appearance acquisition, do actually produce very sharp self-casting shadows. The sharp discontinuity in shadows produce situations like Fig.14 which is actually difficult to remove completely when creating the texture maps, specially with only a few images, leaving some undesirable halos in the texture. However this is not a practical issue since most sources, including the sun on a clear day, are actually extensive to some degree. There is also normally some significant indirect illumination from the environment or other parts of the object itself. Best results are actually achieved in a partially clouded day.
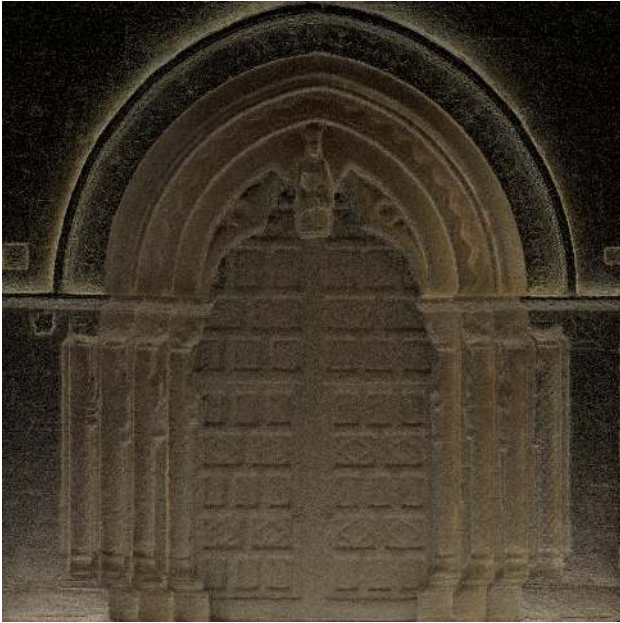
Figure 13: Indirect self-illumination showing the influence of the second bounce of the light in he model, where close surfaces of the own model cast light onto nearby surfaces.
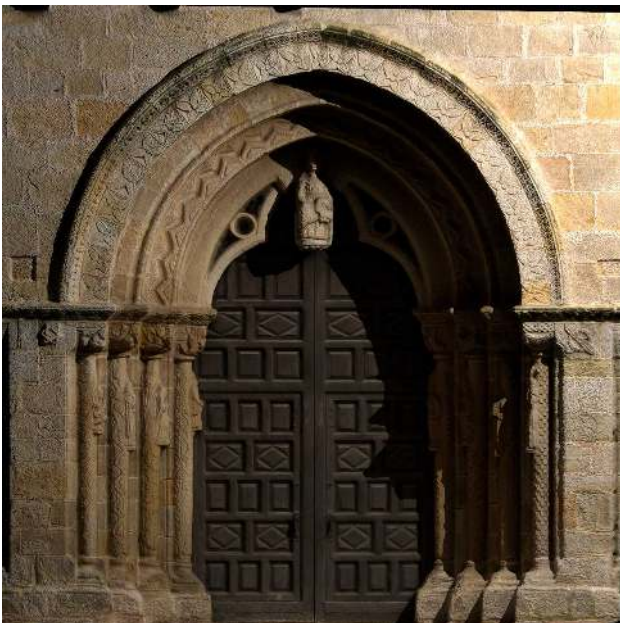

Figure 14: A synthetic render where there is a single point light source, to illustrate the hardness of the shadows when there is no significant environmental indirect light, in contrast with Fig.3. This pattern may occur at night with an artificial spot beam.

layers contain separated information about the depth of the geometry, surface normal fine details and other illumination-dependent parameters such as direct and indirect illumination, or projected self-shadows.

Figure 15: One of the original photographs (up) of the door of Santiago de Peñalba, Leon, Spain and an synthetic view (down) of the 3D model with acquired appearance, rendered using physically based methods for a different viewpoint and illumination.

The enhanced texturing enables a dynamic appearance response to the light when using realistic render methods, as well as the possibility to add, adjust or fully remove shading and projected shadows at will. This is useful not just for achieving a realistic visualization but also for the production of enhanced drawings, with highlighted edges and volumetric shading.

Some of the intermediate layers, specially those containing depth and shading information can be used also for alternative representations providing the look and feel of artistic rendition of volumetric drawings. By processing, including edge detection at different levels, filtering and adjusting relative weights of the various bitmaps, it is possible to generate automatic compositions that try to emulate the process that an artist would do interpreting volumes, shading, and projections of shadows, also standing out further those edges with stronger changes in relief and depth. The result, generated automatically from the re-illuminable orthoimage is show in Figs.12 and 17, and can be used as alternative representations with improved volumetric interpretation.

3. CONCLUSIONS

We have demonstrated a practical methodology for appearance acquisition which was previously introduced in (Martos and Ruiz, 2013), and is specifically applied here for the production of reilluminable architectural orthoimages. It is suitable for outdoor situations, where the illumination can be variable and uncontrolled. Actually the naturally occurring changes in light among different images are desired and exploited to produce and enhanced multilayer color texture, not limited just to a fixed RGB map. These

Finally the whole documentation methodology for outdoor items was demonstrated to be quick, flexible and practical, basically requiring just a digital camera, while still taking advantage of natural or induced variations in environmental light. In both case-studies, *La Oliva* and *Santiago de Peñalba*, photographs were

Figure 16: Door of Santiago de Peñalba, Leon, Spain. Detail of the orthoimage(down) with slightly different slanted illumination to highlight relief details.
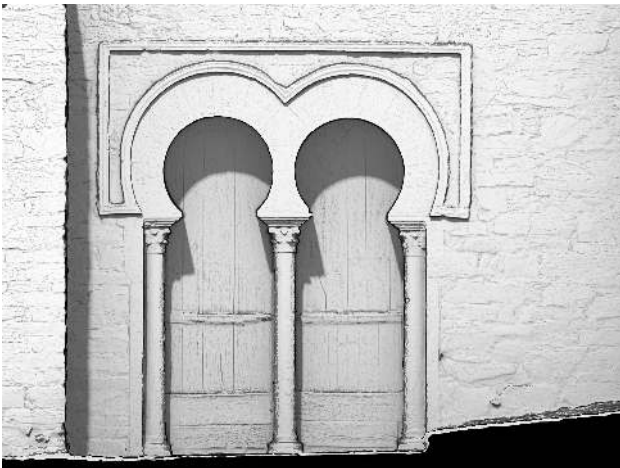

Figure 17: Automatic volumetric drawing generated by processing and composing some of the multiple layers with different edge detection methods (base RGB color information is removed).

taken under variable yet uncontrolled illumination conditions (ie, different times of the day) in a few hours or event minutes, without much preparation and without significantly increasing field work when properly planned. As a result the surface appearance could be still acquired and rendered successfully with complete flexibility to re-illuminate the object at convenience.

REFERENCES

Basri, R., Jacobs, D. and Kemelmacher, I., 2007. Photometric stereo with general, unknown lighting. International Journal of Computer Vision 72(3), pp. 239–257.

Callieri, M., Cignoni, P., Corsini, M. and Scopigno, R., 2008. Masked photo blending: Mapping dense photographic data set on high-resolution sampled 3d models. Computers & Graphics 32(4), pp. 464–473.

Chen, T., Taplin, L. A. and Berns, R. S., 2011. Munsell color science laboratory artist material database brdf fitting. Munsell Color Science Laboratory Technical Report, www.art-si.org.

Furukawa, Y. and Ponce, J., 2010. Accurate, dense, and robust multi-view stereopsis. IEEE Trans. on Pattern Analysis and Machine Intelligence 32(8), pp. 1362–1376.

Haindl, M. and Filip, J., 2013. Visual Texture - Accurate Material Appearance Measurement, Representation and Modeling. Advances in Computer Vision and Pattern Recognition, Springer.

Klein, R. and Schwartz, C., 2010. 3D-Coform. 2nd year report. [Online; accessed 06-May-2013].

Malzbender, T., Gelb, D. and Wolters, H., 2001. Polynomial texture maps. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, ACM, pp. 519–528.

Martos, A. and Ruiz, B., 2013. Realistic virtual reproductions. image-based modelling of geometry and appearance. In: Digital Heritage International Congress (DigitalHeritage), 2013, Vol. 1, IEEE, pp. 127–134.

Palma, G., Callieri, M., Dellepiane, M. and Scopigno, R., 2012. A statistical method for SVBRDF approximation from video sequences in general lighting conditions. Eurographics Symposium on Rendering 31(4), pp. 1491–1500.

Schwartz, C. and Klein, R., 2012. Acquisition and presentation of virtual surrogates for cultural heritage artefacts. In: EVA 2012 Berlin, Gesellschaft zur Förderung angewandter Informatik e.V., Volmerstraße 3, 12489 Berlin, pp. 50–57.

Tunwattanapong, B., Fyffe, G., Graham, P., Busch, J., Yu, X., Ghosh, A. and Debevec, P., 2013. Acquiring reflectance and shape from continuous spherical harmonic illumination. ACM Transactions on Graphics (TOG) 32(4), pp. 109.