EUROGRAPHICS Workshop on Graphics and Cultural Heritage (2016)
C. E. Catalano and L. De Luca (Editors)

A framework for compact and improved panoramic VR dissemination

B. Fanini & E. d'Annibale

CNR ITABC, Rome, Italy

Abstract

Panoramic capture devices in Cultural Heritage are becoming widely available to consumer market, also due to comfortable interactive online dissemination and to the growth of VR segment. VR fruition through an HMD although, requires a virtual 3D representation to provide consistency in terms of experience, scale and spatial perception, overcoming limitations of standard approaches in orientation+positional HMD tracking model. However, modeling of 3D scenes and especially optimization of acquired dataset, are often time-consuming tasks: these are further stressed when dealing with latency-free demands of latest HMDs. In this paper, we propose a novel framework for panoramic acquisition and an improved data model for VR dissemination: spherical panoramas, omnidirectional depth-maps and semantic annotations are encoded into a compact, coherent representation that suits modern HMD needs and low-cost VR devices. We describe advantages of our approach in terms of acquisition pipeline, presence and depth perception in HMDs fruition, discussing also visualization efficiency in online contexts. We present a few case studies where we applied the methodology and the workflow we adopted, comparing results. We discuss integration of existing desktop toolkits into the pipeline, dissemination capabilities through recent WebVR API and framework advantages for immersive VR panoramic video streaming.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual reality H.5.2 [INFORMATION INTERFACES AND PRESENTATION]: User Interfaces—I.4.10 [IMAGE PROCESSING AND COMPUTER VISION]: Image Representation—Multidimensional I.4.1 [IMAGE PROCESSING AND COMPUTER VISION]: Digitization and Image Capture—I.3.3 [Computer Graphics]: Picture/Image Generation—Bitmap and framebuffer operations

1. Introduction

In recent years, panoramic capture devices are becoming cheaper and more accessible to consumer market, making them very appetible for fast spherical acquisition and interactive dissemination applied to Cultural Heritage field and to other sectors as well. Spherical panoramas nowadays, have several purposes, for instance exploited as directional maps, often used in computer graphics industry for local image-based illumination, reflection effects and much more. They are especially appreciated in immersive applications for their ability to *"fill"* the user peripheral vision and the creation of an engaging interactive context, besides their speed and convenience for real-time purposes. Within physical setups, in order to create an immersive environment for the user, general approaches employ multiple screens or different projections. In general, these solutions introduce efforts into edge blending to create a seamless panoramic experience and more in general, they have major impact on software/hardware requirements. Recent introduction of consumer-level Head-Mounted Displays (e.g.: Oculus Rift, HTC Vive, Samsung Gear VR, and others) can offer cheaper deployment in this specific context to create virtually seamless experiences, by removing any imagery pollution from external real world. A correct VR experience using full degrees of freedom for head orientation model is delivered from a 3D representation of a virtual environment. Since modern HMDs provide full stereoscopic vision and good peripheral ranges (e.g. HTC Vive has a display field-of-view of 110 degrees per eye) in order to offer a smooth interactive experience, content providers face several challenges, including dual rendering, culling performance due to larger FOVs, assets optimization and more. Recent growth of WebVR API popularity (`https://webvr.info/`) is very appealing in this context for VR fruition through a standard browser, although additional issues are raised due to 3D content optimization and streaming. On the other hand, panoramic images and videos, offer good environment approximations using simple and compact data representation, with generally fast workflow. For immersive panoramas on HMDs, there are in general two mainstream approaches: projection of the equirectangular image onto a fixed radius sphere with virtual camera placed in the exact center, or by means of stereoscopic panoramic pairs. The first approach offers good freedom in terms of

DOI: 10.2312/gch.20161380

head orientation model (full *yaw-pitch-roll*) although the perceived depth is *"flat"*, due to the monoscopic nature of the panoramic image and the constant egocentric distance (the distance from user point of view to a target). The second approach with stereoscopic pairs works correctly in two situations: (a) user has no freedom over the virtual camera orientation and (b) user is not allowed to apply roll motion (rotation around the view-direction vector) due to the implied up-direction in the pair [Bou06] and limited in other head rotations. In the latter case, motion sickness or eye strain are introduced due to mismatching up directions, causing discomfort. Within the framework presented in this paper, we describe optimized workflow, data model and dissemination approaches that combine panoramic images and videos with depth information to overcome limitations in head orientation and position, restoring correct scale and depth perception in VR fruition. We enrich the data model with efficient encoding and representation of semantic annotation maps, while maintaining the interactive advantages of omnidirectional approaches.

2. Related work

Depth acquisition techniques have already been used in spherical projection context to enhance standard equirectangular panoramas in previous work. An example is the DEP model [BPS04] where both color and geometry are sampled by means of acquisition device. Such approach is employed in order to overcome the efforts of manual 3D modeling for complex scenes. Other works include combination of captured panoramas and 3D elements [FZV13] for consistent illumination models: this is again obtained by means of panoramic depth-maps generated from modeled geometries, associated with equirectangular images. Omnidirectional depth-map in this case is also taken into account for light scattering and reflections to obtain photo-realistic rendering. Previous research already focused efforts on compression and encoding of depth information in stereo rendering [FHF03] [Feh04] and recent work on computation of omnidirectional depth-maps from cheap 360 acquisition cameras [BTH15]. Regarding egocentric distances, recent studies also focus on the importance of depth perception in modern HMDs and cinematic VR [TBL*16], investigation of distance compression [dCAdS15] and factors that may influence user depth perception [RVH13].

3. The proposed framework

In this section, we describe our general approach, methodologies and workflow starting from panoramic acquisition to data encoding for interactive VR dissemination, including online scenarios. In the first phase, we describe the panoramic acquisition pipeline, including metric accuracy and photorealistic representation resulting from the integration and the implementation of different techniques and devices. In the second phase we describe the encoding of multi-layered omnidirectional information into a compact data representation for acquired spherical panorama, depth-map and semantic annotation map. We introduce the *DPF* (Depth Panoramic Frame) as carrier of omnidirectional image-based information (panoramic images and videos) for VR, in both online and local contexts, without the need for geometrical 3D assets streaming. Depth encoding in proposed data model is discussed: specifically distribution optimizations in order to maximize accuracy, taking into account egocentric distance perception on HMDs. An important task within the proposed framework is how to produce such equirectangular depth-maps: there are multiple approaches that will be further described at the end of next section, focusing on acquisition and generation of image-based depth information. We describe results in terms of workflow boost and dissemination, with constant and highly interactive frame-rates provided by our approach, HMD rendering and its portability on low-end headsets. The annotation layer and its encoding are finally presented, including advantages in terms of workflow and data optimization compared to 3D geometrical descriptors for semantic enrichment [SSRS12].

3.1. Acquisition and 3D workflow

The proposed acquisition pipeline within the framework is a result of the integration and the implementation of different techniques, aiming at the achievement of high metric accuracy and photorealistic representation. Approaches to 3D documentation and survey can be more or less direct, effective and affordable in terms of economy and professional use. The workflow presented in this section is based on a broader research background [Fan07] studying how multiple spherical panoramas can be used as low-cost, accurate, affordable and complete source of 3D documentation [d'A11]. Actually due to integrated tools for Structure From Motion and Dense Stereo Matching (Agisoft Photoscan `http://www.agisoft.com/`) photogrammetry procedures are mostly automated and provide good efficiency in terms of accuracy and robustness.

3.1.1. Photo acquisition and data orientation

Within the presented framework, there are mainly two different strategies for panoramic acquisition: a "classic" slower approach and a quicker one. They are performed by two different devices: a reflex camera using a tripod with panoramic head and 360 RICOH Theta S camera (`https://theta360.com`), a cheap, comsumer-level *1-shot* panoramic device. The first approach guarantees good

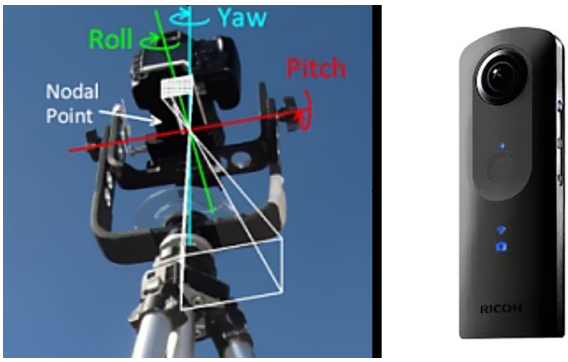

Figure 1: *Acquisition devices*

stability, crucial when it's necessary to work with long exposure times because of poor illumination. A sharp image is decisive when carrying out a SfM procedure: a bad acquisition can in fact lead to no results or missing parts. The advantages of such approach

(camera+panoramic head) are (1) the final resolution of the images, (2) a good control of exposure parameters and (3) color balance. Finally, a stitching software is necessary to create the spherical panorama. Common software tools can be used for latter task, such as *PTGui*, *Hugin* or *Autopano* using state-of-the-art stitching methods [SS97] [Sze06]. The second approach using the RICOH Theta S camera has two limits: low resolution (5376×2688) and exposure/color control. Although in a single shot is possible to produce a spherical panorama, since the internal software is responsible for the stitching. In order to locate and transform the spherical images in the virtual environment, a photogrammetric approach is used: the orientation of images uses corresponding points and coordinates of a few targets (eventually used to geo-reference the model). The feature detection procedure can be carried out manually for instance using *Sphera* Software [Fan07] - generally requiring a low number of points for orientation - while in Photoscan, both orientation and reconstruction are automatic.

3.1.2. Acquisition and production of omnidirectional depth information

In order to generate reliable omnidirectional depth information from a given position and orientation in a target space, good metric accuracy and detail are often required. There are generally two possible scenarios: depth information is (1) directly acquired by means of a physical device or (2) computed from a virtual, geometrical 3D representation of the target space. The first class include for instance laser scanner devices, that besides their effectiveness and dense accuracy, are often expensive solutions. Furthermore such devices require also additional processing in order to obtain omnidirectional depth information, where automated software tools are not provided out-of-the-box. Some recent techniques like vertical camera displacement applied to cheap 360 cameras, allow computation of omnidirectional depth from a given position, except zenith and nadir [BTH15], although additional digital intervention and processing are necessary. Within the second class, recent approaches [PGG*16], consumer level devices like RICOH Theta and other recent sensor cameras, offer cheap and quickly improving solutions to recreate a 3D representation of the environment. It is quite noticeable from different works that best results in terms of noise, accuracy and scale are obtained in general by considering a virtual geometric representation of the scene, rather than relying on disparity maps computed from a single stereo pair. Our focus is in fact to achieve a *reliable* and seamless omnidirectional depth information in equirectangular space, minimizing error for d.

$$d = D(x, y)$$

In order to obtain such accurate depth information, a solution is to perform the computation early in 3D workflow - from raw point-clouds or draft 3D representations generated using cheap panoramic devices. In our framework in fact, one goal is to remove common slowdowns and bottlenecks in 3D optimization pipeline. A dedicated tool - developed as server-side service and as client application - was deployed in order to validate and support such workflow, thus producing omnidirectional depth-maps starting from acquired point-clouds or mesh-based geometries. The encoding of such collected depth information will be further described in section 3.2.1.

3.2. DPF encoding and dissemination

The DPF (Depth-Panoramic Frame) is a proposed image-based model for compact representation of omnidirectional data, tailored to real-time fruition and remote streaming of depth-enhanced panoramic images and videos. The general idea is to encode color+depth information and semantic description, while maintaining suitable image format for digital intervention, update or enhancement. Each DPF contains orientation and position data of acquired panorama: such information is used by the application to restore and arrange the dataset in a 3D space. VR fruition (using full orientation+positional tracking model) and depth perception through HMDs is in fact improved by recovering an approximation of original 3D environment from a specific position in space. Following sections will describe more in detail depth and semantic omnidirectional representations, presenting advantages in workflow, fruition and finally VR deployment and user interface to interactively query annotation layers.

3.2.1. Depth maps

Depth maps in DPF are encoded and quantized into a 8-bit image, in order to overlay such data with other image-based information, for instance on separate channels (RGB). Such limited storage is although sufficient to restore depth information by taking into account the following considerations:

1. Depth fruition in HMDs and perceived distances by user
2. Required accuracy for a given egocentric distance

For 2 we consider the following scheme, by highlighting 3 different distance segments and their required depth accuracy: (1) Very close objects; (2) medium range distances and (3) far objects. Far objects do not require in general high accuracy due to human stereoscopic perception in nature, while closer items or surfaces require more depth precision. For such reasons, a quadratic distribution has been employed in order to optimize accuracy and depth perception within limited storage (see Figure 2). Stored depth values are normalized in $[0,1]$: an interactive depth range $[min, max]$ into the DPF is finally used to recover original depth from original acquired position at runtime. A completely GPU-based approach can be used (e.g. using vertex shaders) to deform a standard unit sphere (with radius = 1), while maintaining the geometry tessellation (sphere ratio) separated. This can be especially useful in panoramic video contexts, typically requiring major workload on the CPU. Figure 3 shows an acquired panoramic image projected

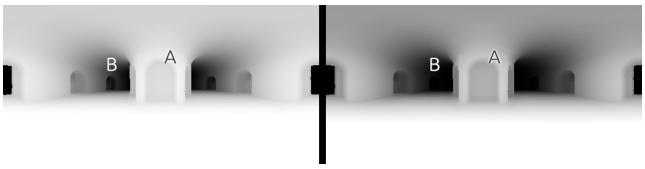

Figure 2: *A comparison between linear depth values (left) and quadratic distribution (right) generated by the service, encoded into an omnidirectional depth-map. The quadratic distribution improves depth accuracy on closer details (A) and sacrifices it on distant details (B)*

onto a standard unit sphere - with constant egocentric distance - compared with a DPF carrying color and depth information previously described in Figure 2, to recover original distances given specific position and orientation. Due to quadratic distribution and interpolation performed on the GPU, it is possible to use low resolutions depth-maps and still retain good results (more details will be presented in section 4).

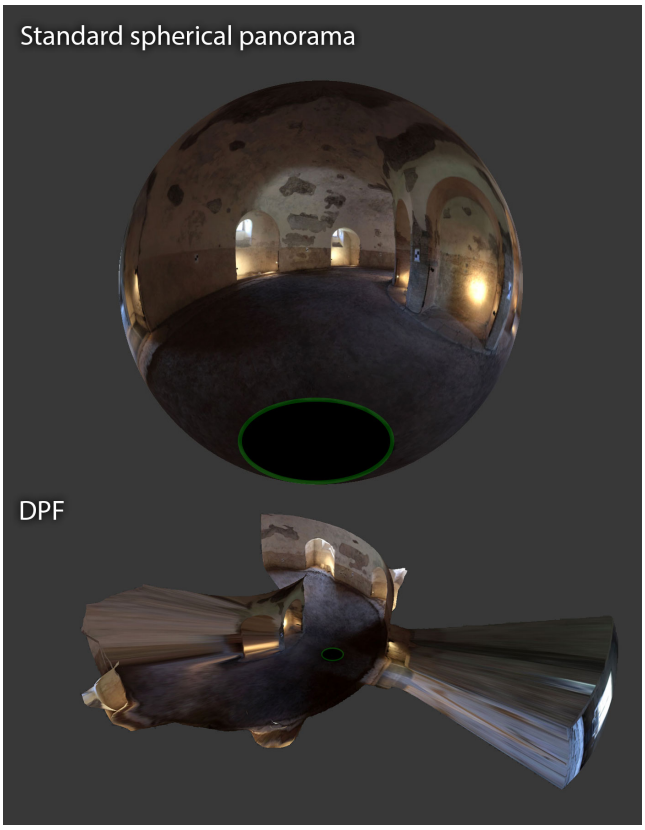

Figure 3: *Panoramic acquisition projected on a unit sphere (top) and with restored depth values encoded into a DPF (bottom) from a specific position (the black disk) and orientation. Panoramic fruition from origin is undistorted in both cases, although the right provides correct depth and scale information*

3.2.2. Semantic annotation

Semantic description should leverage on a consistent spatial encoding, and - from a 3D user interface perspective - interactive interrogation of semantically enriched panoramas should take advantage of depth perception offered by the described approach. Since the image-based approach embodied by DPF model, a natural and suitable description for multiple annotation layers designed for efficient queries, can be encoded as equirectangular data as well. In a virtual 3D environment, a common way to semantically enrich a virtual scene is through geometrical definitions by means of simplified shapes [SSRS12]. For complex scenes, this task can be time consuming, even deploying semi-automated approaches for generation of queryable geometrical shapes. Within omnidirectional images - and thus within the DPF model - the problem is vastly simplified due to the egocentric projection, thus facilitating annotation tasks by digitally performing them on a bi-dimensional dataset. A crucial aspect at this point regarding online transmission is how to encode and optimize such spatial description for several semantic areas. The approach used in the proposed model is a single map encoding multiple boolean masks, each representing an annotated area. Figure 4 shows an example of four equirectangular annotations masks (A, B, C and D) and their automatic combination into the omnidirectional semantic map (bottom left), maintaining the representation suitable for further digital intervention.

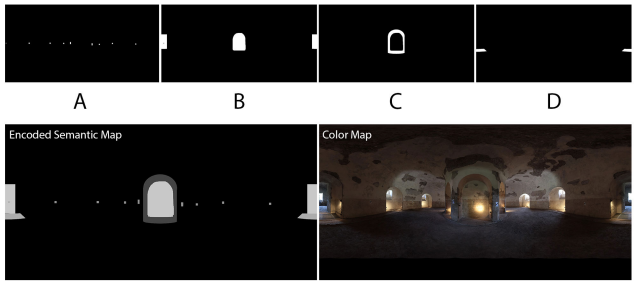

Figure 4: *Four annotation masks (A,B,C and D) encoded into semantic map (bottom left).*

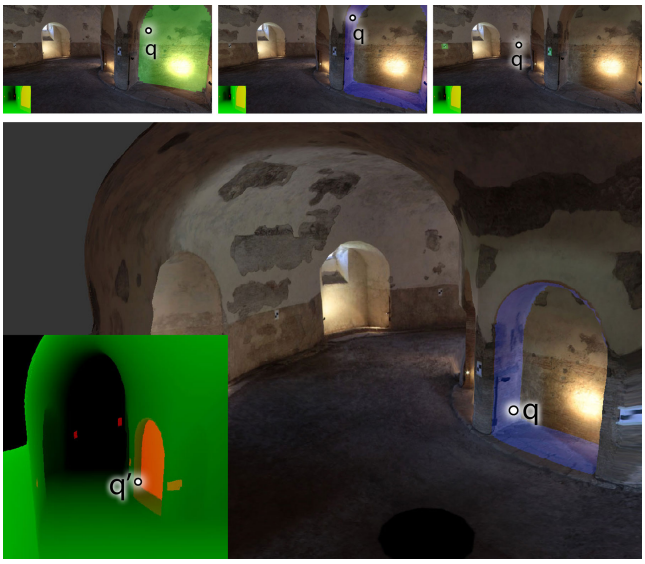

Figure 5: *a screen-space query (q) performed on visible space at different cursor positions (upper row) and mapping it in hidden semantic-space (q') (bottom). In this case the interactive semantic virtual camera encodes both annotations (red channel) and depth information from current position (green channel)*

An annotation function $A(x,y)$ is performed to query such map: the resulting value v, when non-zero, indicates the requested coordinate is annotated. An hashing procedure H is used to encode and decode a specific annotation value during interactive query operations:

$$v = A(x,y) \quad H(v) = h \quad H^{-1}(h) = v$$

Multiple contiguous values in a specific interval $[v_a, v_b]$ (with $v_a < v_b$) collapse into a single hashed annotation code h, thus reducing the maximum annotations areas allowed per DPF: such quantization depends on image or video stream compression for a given scenarios (e.g. lossy streams will employ stronger quantization in H). From a workload perspective, segmentation and generation of queryable annotation masks is greatly simplified, due to the nature of the representation. Common digital painting software and tools can be deployed to easily segment areas or objects, then associate hashed annotation codes with specific content (images, audios, text, movies, etc.) that is presented to the user during interactive fruition. It is important to specify that interactive queries must to be performed in the same 3D context deformed by depth-maps: this is required for a consistent mapping between visible space and semantic space (hidden to the user). Figure 5 shows an implementation of non-VR prototype: mouse position in screen-space is used to query different spots in visible virtual space by mapping semantic space (lower left) - both 3D spaces are transformed by the depth-map.

3.2.3. VR fruition

As previously described, our approach provides real depth perception in a 3D restored space targeting head mounted displays fruition. This approach allows (1) a full *yaw-pitch-roll* orientation model, maintaining stereoscopically correct results due to the 3D nature of the virtual environment approximation; (2) correct scale perception and (2) a limited radius where the user can also take advantage of positional tracking, commonly provided by modern desktop HMDs. The latter radius strongly depends on different factors, mostly including sudden depth variations on the equirectangular panorama and minimum depth range of the DPF: larger values allows larger head translations into the physical space without the user noticing the occluded panoramic areas. Furthermore, as previously introduced, increasing refresh rates of modern HMDs demand robust frame-rates and minimal or absent latency. This strongly impacts common assets optimization tasks, from both a geometrical and a texturing perspective. With DPFs approach, we are providing constant geometry and texture footprints, thus maintaining a constant and smooth experience, even on low-end HMDs, independently from the original 3D asset complexity and its variability, acquired or modeled. Regarding the 3D presentation of annotated content and VR fruition in general, special attention must be given into interaction design in order to preserve a comfortable experience for the user. Best Practices provided by Oculus and previous research [DBB15] [SW11], highlight a few common design guidelines to avoid eyestrain and discomfort that suit the fruition of a DPF. 3D queries in the virtual space are performed from user head position and view direction: this also takes into account additional transformations (translations) provided by positional trackers. As introduced before, such queries do not require 3D intersectors or geometry-dependent operations, thus not impacting framerate at all

HMD device	FoV (degrees)	Resolution (W x H)	Min. Required Equirectangular Width (pixels)
Oculus Rift DK2	100	1.920 x 1.080	~ 3.456
OSVR Razer	100	1.920 x 1.080	~ 3.456
Oculus Rift CV1	110	2.160 x 1.200	~ 3.534
HTC Vive	110	2.160 x 1.200	~ 3.534
Star VR	210	5.120 x 1.440	~ 4.388
Samsung Gear VR	96	2.560 × 1.440	~ 4.800

Figure 6: *Comparison between different HMDs, display resolution and display FOV. A corresponding minimal width resolution for panoramic frame in relation to dFOV and HMD resolution is shown (last column): a* 4096×2048 *equirectangular for instance can be sufficient to exceed most HMDs pixel densities*

and providing a smooth experience. Presenting annotated content should not infringe the golden rule of *"immersion from start to finish"* by occluding the entire user field-of-view, but rather use 3D surfaces at appropriate depth and size to render annotation content. 3D overlays should be also consistent with full yaw-pitch-roll head orientation model and positional tracking (see Figure 7). Highlighted areas should respect the perceived depth (panoramic surface): this is easily accomplished by using for instance GPU shaders to gently apply pulsating effects to the semantically enriched area. More results on VR fruition applied to a few case studies will be presented and discussed in the next section.

4. Case studies and results

In this section we present a few case studies where we applied the proposed framework, including workflow, DPF model and VR fruition through head-mounted display. We compare pipelines in 3D model processing, data size of 3D assets and corresponding DPFs. In order to provide better comparisons, multiple factors should be considered: acquisition device costs, acquisition and data elaboration times. The first case focuses on the inner section of the *Mausoleum of Romulus* in Rome, where a previous work [ACd*14] was carried out with the objective of comparing two different photogrammetric approaches: dense image matching and spherical photogrammetry. The latter approach, including topographic data and 3D assets already produced, provided a perfect testbed for the DPF model.

Regarding the Mausoleum case study, we started from the acquired 8 spherical panoramas, the point-cloud generated from laser scanner and the image-based modeling [d'A11] to compare differences with geometrical meshes. Spherical panoramas were oriented by Multi-Image Spherical Photogrammetry (MISP), a convenient photogrammetric technique particularly suitable for metric recording of architectures [Fan07]. Our goal was specifically to under-

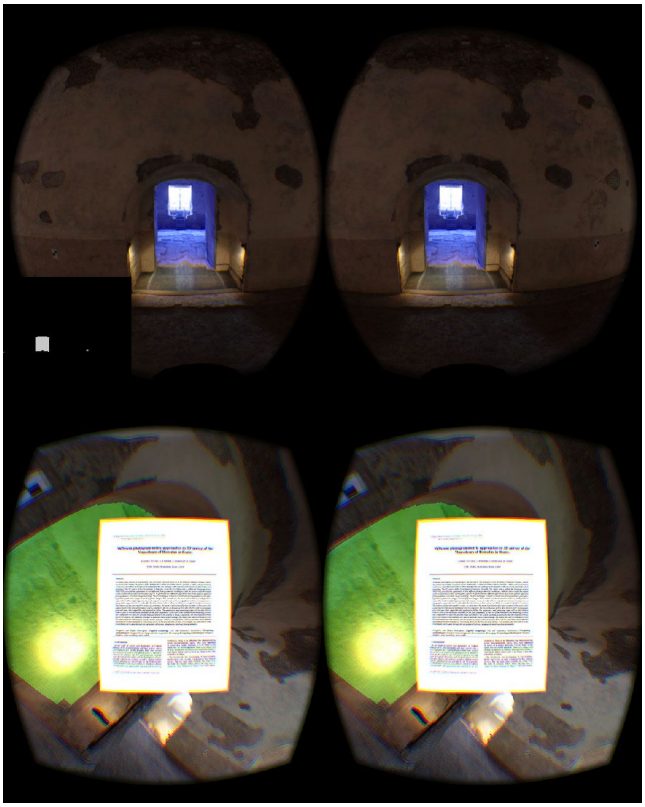

Figure 7: *Highlight of far annotated area in 3D environment (top) and presentation of annotated content in 3D space respecting full yaw-pitch-roll head orientation model)*

stand how early is possible to employ the described framework obtaining good results in terms of performance and accuracy in VR dissemination. Specifically if the DPF model can be deployed in early stages to stream an immersive 3D environment before time-consuming tasks such as simplification and optimization. We tested the omnidirectional depth service with quadratic distribution, on both point-cloud and optimized geometrical mesh using corresponding positions and orientation data. During the tests we used 512×256 and 256×128 resolutions with PNG lossless format for all the 8 oriented panoramas, in order to compare depth and scale restoration results within real-time fruition. Results showed that low-resolution depth-maps (256×128) were sufficient to restore correct depth with good detail. Furthermore, computed maps from point-cloud were very close as result compared with those operating on mesh-based asset (see Figure 10), suggesting that we could start to employ the depth service and DPF in early pipeline stages. The implemented service for depth computation employs an optimized tracing procedure on geometrical meshes, while for point-clouds it gathers incoming points distances and applies interpolation on resulting data. In both cases: (a) depth range $[min, max]$ is extracted from gathered data $D(x,y)$, (b) values are normalized and (c) a selected distribution is applied, as described in section 3.2.1. The two DPFs shown in Figure 11 indicate good accuracy in depth and scale restoration, considering the limited depth-map resolution

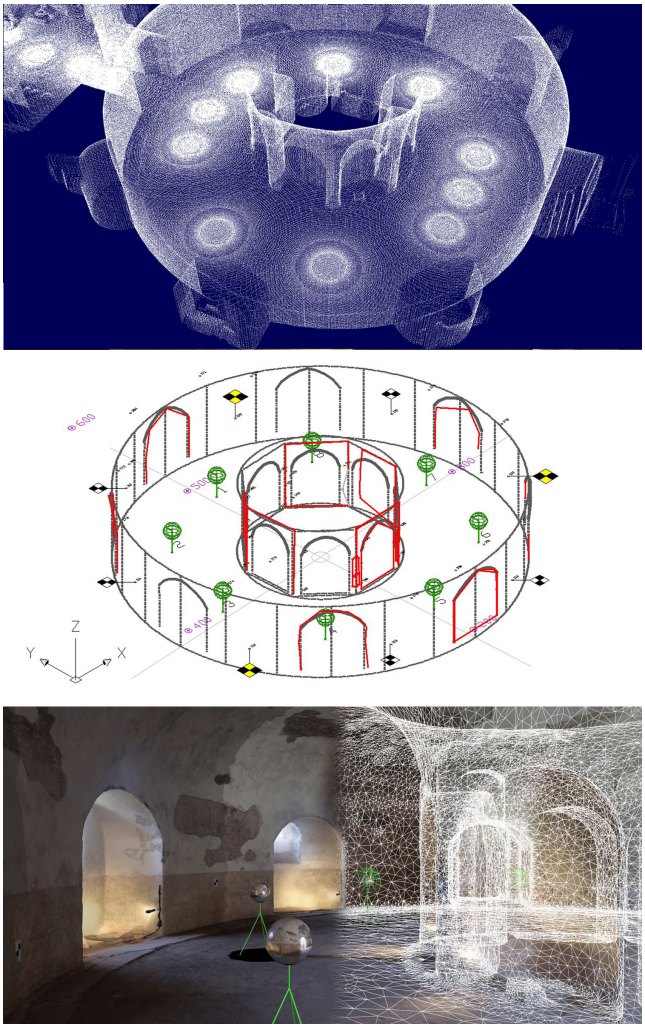

Figure 8: *Starting point-cloud from laser scanner (top); oriented panoramas (center); 3D model obtained by mesh reconstruction, optimization and panoramic texture projection (bottom)*

Nstaz	X	Y	Z	hstr	teta0-rad	alfax-rad	alfay-rad
1	0,14446	47,47731	-1,55160	1,5	6,22325	-0,00333	0,00062
2	-5,51362	45,38706	-1,52457	1,5	5,47206	0,00285	0,00069
3	-7,65191	39,18847	-1,49363	1,5	4,67783	-0,00149	-0,00018
4	-5,40889	33,21638	-1,50947	1,5	3,91385	-0,00191	0,00123
5	1,03690	30,97371	-1,50645	1,5	3,08981	0,00254	0,00128
6	7,45930	33,82308	-1,53500	1,5	2,30457	-0,00300	-0,00148
7	9,38338	39,91188	-1,57813	1,5	1,52363	0,00247	-0,00046
8	6,37139	46,06691	-1,55526	1,5	0,73719	-0,00094	-0,00204

Figure 9: *Orientation data by MISP*

used (256×128). It is also noticeable how surfaces closer to acquisition points are more precise in terms of distance due to quadratic distribution.

In another case study involving Scrovegni Chapel in Padua (Italy), only 4 panoramic acquisitions were used and a quicker pipeline was employed, thanks to automatic procedures for image

Figure 10: *Panorama #3: top row shows differences between computed omnidirectional quadratic depth-map from point-cloud (left) and from mesh-based model (right). Bottom is panoramic color acquisition as reference*

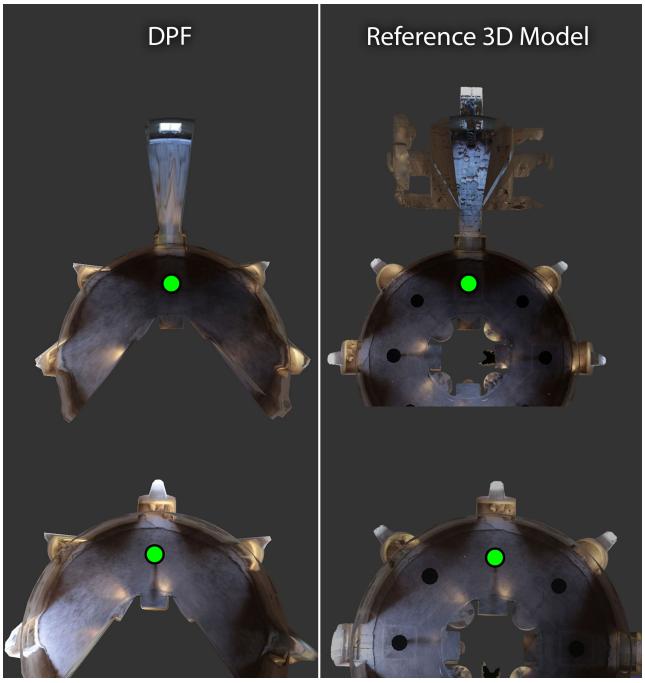

Figure 11: *a comparison between two DPFs (left column) with acquisition points (green) of panoramas #1 (top) and #3 (bottom) and reference 3D model (right column) with processed geometry + textures*

orientation offered by Agisoft Photoscan. The overall pipeline took approximatively: (A) ≈1 hour for panoramic acquisition, (B) ≈2 for stitching and creation of four 24576 × 12288 resolution panoramas, (C) ≈ 1 hour for orientation in Photoscan (see see Figure 12) , (D) ≈2 hours for point-cloud generation and (E) ≈4 minutes for 4 omnidirectional depth-maps, automatically produced by the service. In conclusion, these tasks were carried out within 1 working day - see Figure 13.

#	x	y	z	Omega	Phi	Kappa
[Pano-01]	5,844693	25,559522	-1,9932833	90,37254	-0,31652	-1,166140273
[Pano-02]	5,870517	12,675974	-2,0453164	90,022437	-0,36995	-0,386736577
[Pano-03]	5,904258	4,2694939	-1,9934434	90,47231	0,068666	-0,338989551
[Pano-04]	5,884034	31,577536	-1,8636512	90,073819	-0,54257	-0,163614098

Figure 12: *Orientation data in Scrovegni Chapel (local coordinate system)*

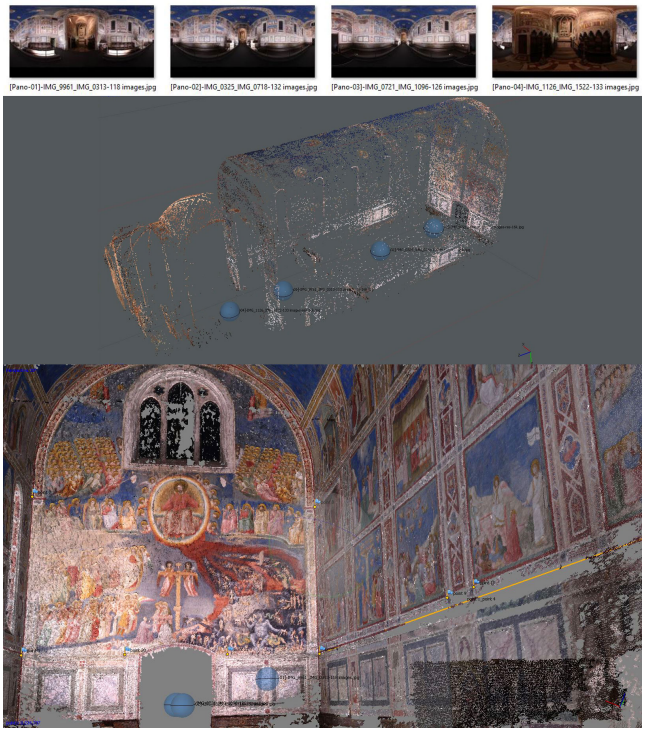

Figure 13: *Panoramic images (top), tie-points of oriented panoramas and dense cloud with 8.234.737 points (bottom)*

While panoramic color information in this case required more storage (≈700Kb at 4096 × 2048) in order to maintain good chromatic properties, the omnidirectional depth-map required ≈8Kb (at 256 × 128 resolution) to recover original egocentric depths and scale of the 3D environment during HMD fruition. Color information size can be easily fine-tuned through JPEG compression algorithms provided by common graphics editors, depending on the dataset and color accuracy we want to stream to the VR device. Another result for this use case was assessing the fast annotation pipeline, that involved common 2D painting tools like GIMP (`https://www.gimp.org/`) to semantically enrich the cycle of frescoes and framed scenes (see Figure 15). Sample tasks were carried out in order to attach full-resolution paintings to corresponding areas by means of 3D rendered overlays (see previous section on VR fruition) shown upon VR spatial interrogation.

The third case study is the *"Sala Ottagona"*, within the Domus

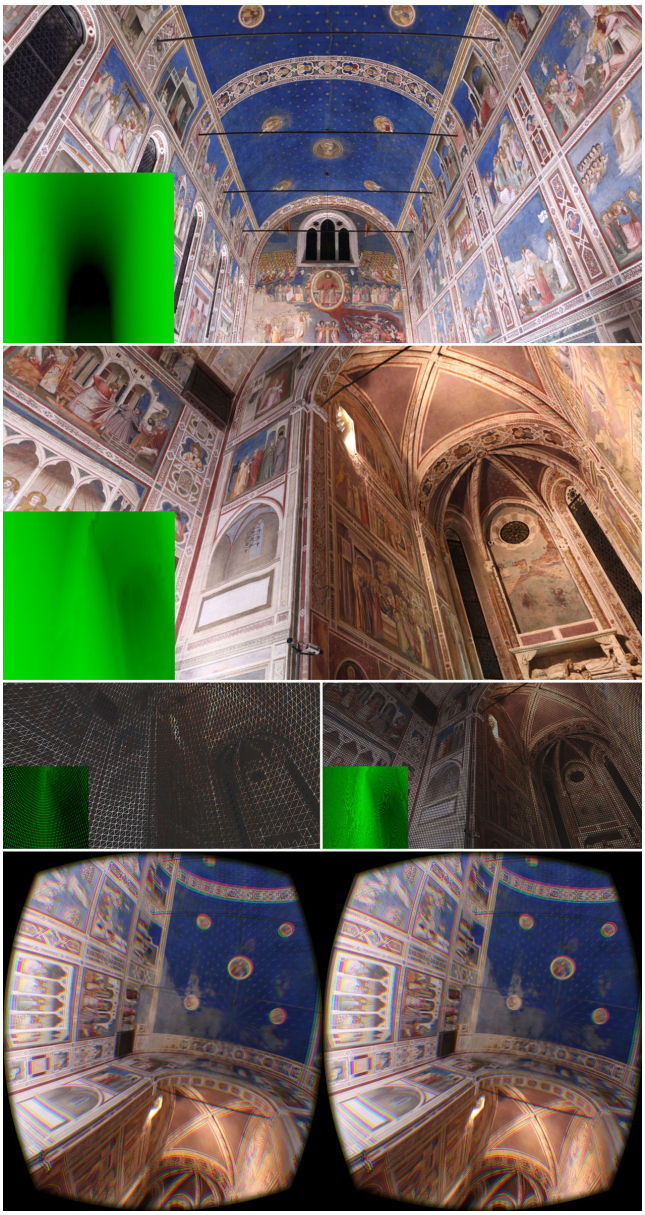

Figure 14: *From top to bottom: Interactive rendering using a DPF, with reference depth camera (in green); sphere tessellation to control interactively the accuracy of 3D approximation; VR fruition with correct scale and depth perception*

Aurea in Rome: in this context our goal was to test out two different camera devices for panoramic acquisition and generation of corresponding point-clouds. The first was carried out using a reflex camera on a tripod with panoramic head and the second using the 360 RICOH Theta S device. These resulted in following datasets:

1. 15 panoramic images (24576 × 12288 resolution each) generated from multiple shots: 19.000.000 points
2. 25 one-shot panoramic images (5376 × 2688 resolution each): 4.400.000 points

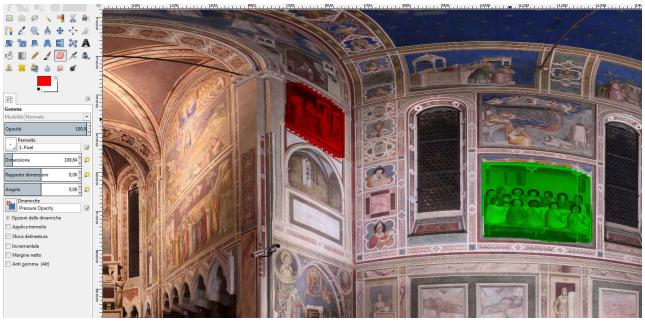

Figure 15: *after panorama generation, a common 2D graphics editor (GIMP) can be employed to digitally paint annotations on top of color information, to produce annotation masks and semantically enrich the panorama*

In this case the overall acquisition took ≈2 hours using the first approach, while the second approach took a few minutes. For both, Agisoft Photoscan was employed to orient panoramic images and generate corresponding point-clouds. The second approach was indeed faster in terms of overall acquisition times, although the first gave us more accurate and complete datasets. The two resulting point-clouds were used to generate egocentric depth-maps for DPF representations: in this case performance tests were carried out on both desktop and online fruition through WebGL. The latter was implemented as open-source library leveraging on WebVR API and the described data model. The developed testbed demonstrated fluid and constant framerates during fruition on Oculus Rift DK2 and CV1 using latest Chromium and Firefox browsers (see Figure 17), recording constant and maximum framerate available on tested HMDs.

Figure 16: *Comparison between tie-points (top) and between dense point-clouds (bottom)*

Finally for the last case we took advantage once again of the Mausoleum dataset: our goal was to test video streaming approach applied to DPF representation model. This time we interactively rendered the 3D model using a panoramic (omnidirectional) virtual camera and streamed to the WebVR client. We created a frame layout suitable for client-side GPU decoding of multi-layered information contained in a DPF (see Figure 18), with depth and semantic

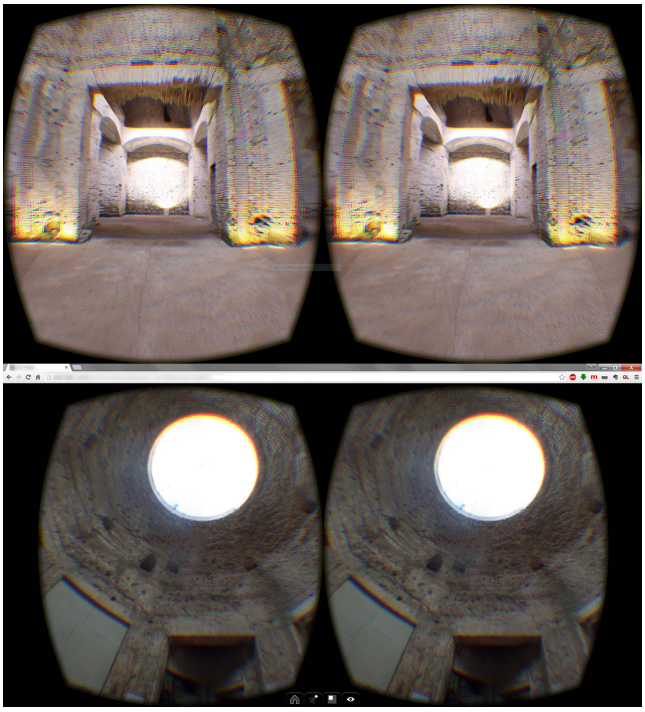

Figure 17: *VR Fruition of a DPF in desktop application (top) and VR fruition through WebVR API, in a browser (bottom)*

maps stretched for border interpolation needs. It was thus possible to stream a complete virtual tour inside the Mausoleum, carrying depth, color and dynamic semantic information in a compact manner at constant frame-rate, allowing the user to really perceive full scale and depth of the 3D environment through the HMD, without actually providing or streaming the 3D model geometry to final client.

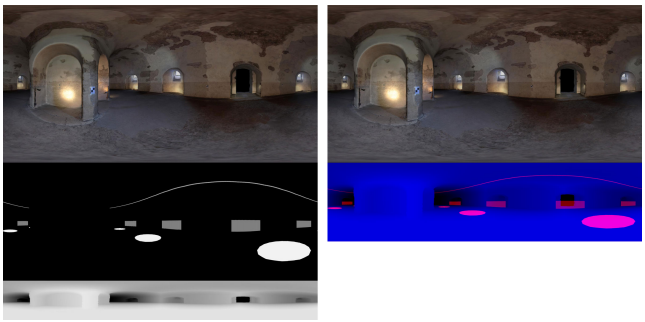

Figure 18: *Two tested layouts to encode a DPF video: panoramic movie, omnidirectional semantic map and depth map (left). On the right, another layout with the same omnidirectional information overlayed into R and B channels*

5. Conclusion and future work

We presented a framework aimed at improving panoramic pipeline in Cultural Heritage, starting from fast acquisition by means of consumer-level devices to a proposed image-based data model (DPF) for enhanced dissemination on VR devices. We define the DPF model as carrier of multi-dimensional panoramic data (images and videos) including (a) depth information to overcome standard limitations in HMD head orientation and positional tracking and (b) semantic omnidirectional information for interactive interrogation. Enconding approaches of such egocentric maps, targeting local deployment and online dissemination, are described: specifically how depth information and semantic enrichment are exploited to restore correct scale and depth perception in HMD fruition, enabling sense of presence. From an immersive VR perspective we discuss efficiency of DPFs, accomodating high framerate demands of modern HMDS and portability of such model on low-end headsets and browsers via WebVR API. Best practices developed within the model are presented in order to provide a comfortable, consistent and smooth experience during VR fruition, discussed in sections 3.2.3 and 4. Improved acquisition workflow of panoramic data is depicted by means of comsumer-level 360 devices and subsequent generation of pointclouds, where the DPF framework discussed in section 3.2 can be employed in early stages before potentially time-consuming tasks on mesh generation and optimization. We finally prove the framework with a series of practical case studies, describing different approaches and results obtained. Within video application, a layout was tested to stream an omnidirectional capture of a virtual tour inside a 3D asset, demonstrating how temporal DPFs can be deployed to allow fluid perception of scale and depth, abstracting from geometrical complexity of original content. Since a DPF encapsulates also position and orientation in space, is already possible to stream multiple DPFs in order to recover different portions of a 3D scene, although this requires masking (via discard maps) computed from inter-occlusions and advanced interpolations while traveling from one DPF to another. Future research will focus on live video streaming using existing toolkits, open-source 3D software (e.g. Blender) and game engines (e.g.: Unreal Engine 4, Unity, etc.) by employing the described model to broadcast in a compact form large and complex 3D environments on high-end HMDs and low-end VR devices, maintaining robust and constant framerates. Another research branch in omnidirectional video streaming will investigate use of branching narratives [KW10] and use of semantic enrichment over temporal axis, for instance using dynamic annotations that change, fade or appear over time.

References

[ACd*14] ADAMI A., CERATO I., D'ANNIBALE E., DEMETRESCU E., FERDANI D.: Different photogrammetric approaches to 3d survey of the mausoleum of romulus in rome. In *Eurographics Workshop on Graphics and Cultural Heritage* (2014), The Eurographics Association, pp. 19–28. 5

[Bou06] BOURKE P.: Synthetic stereoscopic panoramic images. In *Interactive Technologies and Sociotechnical Systems*. Springer, 2006, pp. 147–155. 2

[BPS04] BAHMUTOV G., POPESCU V., SACKS E.: Depth enhanced panoramas. In *Proceedings of the conference on Visualization'04* (2004), IEEE Computer Society, pp. 598–11. 2

[BTH15] BODINGTON D., THATTE J., HU M.: Rendering of stereoscopic 360 views from spherical image pairs. 2, 3

[d'A11] D'ANNIBALE E.: Image based modeling from spherical photogrammetry and structure for motion. the case of the treasury, nabatean architecture in petra. *Geoinformatics FCE CTU 6* (2011), 62–73. 2, 5

[DBB15] DAVIS B. A., BRYLA K., BENTON P. A.: *Oculus Rift in Action*. Manning Publications Company, 2015. 5

[dCAdS15] DE C S., ALYSON M., DOS SANTOS S. R.: Investigating the distance compression on virtual environments by comparing visualization devices. In *Virtual and Augmented Reality (SVR), 2015 XVII Symposium on* (2015), IEEE, pp. 33–41. 2

[Fan07] FANGI G.: The multi-image spherical panoramas as a tool for architectural survey. *CIPA HERITAGE DOCUMENTATION* (2007), 21. 2, 3, 5

[Feh04] FEHN C.: Depth-image-based rendering (dibr), compression, and transmission for a new approach on 3d-tv. In *Electronic Imaging 2004* (2004), International Society for Optics and Photonics, pp. 93–104. 2

[FHF03] FLACK J., HARMAN P. V., FOX S.: Low-bandwidth stereoscopic image encoding and transmission. In *Electronic Imaging 2003* (2003), International Society for Optics and Photonics, pp. 206–214. 2

[FZV13] FELINTO D. Q., ZANG A. R., VELHO L.: Production framework for full panoramic scenes with photorealistic augmented reality. *CLEI Electronic Journal 16*, 3 (2013), 8–8. 2

[KW10] KWIATEK K., WOOLNER M.: Transporting the viewer into a 360 heritage story: Panoramic interactive narrative presented on a wrap-around screen. In *Virtual Systems and Multimedia (VSMM), 2010 16th International Conference on* (2010), IEEE, pp. 234–241. 9

[PGG*16] PINTORE G., GARRO V., GANOVELLI F., AGUS M., GOBBETTI E.: Omnidirectional image capture on mobile devices for fast automatic generation of 2.5D indoor maps. In *Proc. IEEE Winter Conference on Applications of Computer Vision (WACV)* (February 2016). To appear. URL: http://vic.crs4.it/vic/cgi-bin/bib-page.cgi?id='Pintore:2016:OIC'. 3

[RVH13] RENNER R. S., VELICHKOVSKY B. M., HELMERT J. R.: The perception of egocentric distances in virtual environments-a review. *ACM Computing Surveys (CSUR) 46*, 2 (2013), 23. 2

[SS97] SZELISKI R., SHUM H.-Y.: Creating full view panoramic image mosaics and environment maps. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques* (1997), ACM Press/Addison-Wesley Publishing Co., pp. 251–258. 3

[SSRS12] SERNA S. P., SCHMEDT H., RITZ M., STORK A.: Interactive semantic enrichment of 3d cultural heritage collections. In *VAST* (2012), pp. 33–40. 2, 4

[SW11] STUERZLINGER W., WINGRAVE C. A.: *The value of constraints for 3D user interfaces*. Springer, 2011. 5

[Sze06] SZELISKI R.: Image alignment and stitching: A tutorial. *Foundations and Trends® in Computer Graphics and Vision 2*, 1 (2006), 1–104. 3

[TBL*16] THATTE J., BOIN J.-B., LAKSHMAN H., WETZSTEIN G., GIROD B.: Depth augmented stereo panorama for cinematic virtual reality with focus cues. In *2016 IEEE International Conference on Image Processing (ICIP)* (2016), IEEE, pp. 1569–1573. 2