Encoding VR sessions: image-based techniques to record and inspect immersive experiences

Bruno Fanini
VHLab, CNR ITABC
Dept. of Computer Science, Sapienza University
Rome, Italy
bruno.fanini@gmail.com

Luigi Cinque
Dept. of Computer Science, Sapienza University
Rome, Italy
cinque@di.uniroma1.it

***Abstract*—Recording immersive VR sessions performed in virtual environments through HMDs - during public exhibits or permanent installations in a museum - may offer valuable insights on visitors propensities, physical workload, spatial affordances, focus and locomotion data. A first challenge raises from storage, where large amount of casual visitors need to be recorded on a daily basis, analyzed and possibly streamed to remote profes-sionals. For each VR session in fact, size of collected dataset may grow very quickly. Another issue is related to the visual inspection of such data, providing a team of experts tools that allow to revisit the whole history of a single or multiple sessions. We propose, formalize and discuss image-based encoding techniques and quantized signals for compact and light ”musical scores” that allow both common offline 2D image processing and interactive manipulation on modern desktop and mobile GPUs. We describe compact quantization methods and flexible volumetric observers, including location ignition and signatures. The paper discusses obtained results for different case studies where we employed such encodings, including desktop VR, WebVR and streaming of recorded sessions from cheap hardware.**

Keywords—Virtual reality, Data collection, Image analysis, Visualization, Web services

I. Introduction

Within VR deployment and user locomotion, 3D virtual environments targeting Cultural Heritage created for standard or non-stereoscopic visualization generally require large re-adaption for immersive experiences consumed in public exhibits. Such task can present bottlenecks and efforts in terms of modeling and optimization workflow. Furthermore, hard timing constraints within a public expo require special care regarding the user session and its optimization for valuable content. In this sense, we often need to understand spatial affordances and user locomotion in order to improve the overall immersive experience. An interesting discussion is also related to the existence - or the automatic generation - of volumetric ”signatures” for a given VE, synthesized by single or multiple user sessions. In order to understand user sessions from quantitative and qualitative perspectives, recording and collecting per-user data are strongly required operations. Regarding user sessions such task may present several challenges in terms of storage (e.g. large datasets), playback visualization and analysis. Furthermore, within prolonged or permanent exhibits, a remote team of experts may be interested into visualization and analysis of user sessions during the running exhibit itself. The storage and data streaming of recorded users sessions represent a bigger challenge if we deal with deployment of immersive experiences online (WebVR) due to harder constraints given by both bandwidth and 3D visualization capabilities. The contribution of this work specifically targets compact image-based encoding techniques for immersive user sessions, taking into account modern hardware capabilities, storage, portability, compression and online streaming. We introduce and discuss special encoding approaches in terms of volumetric record, data accuracy and transmission, spatial features specifically conceived for immersive VR experiences using HMDs. We discuss advantages of such data encodings in terms of:

- Storage and visualization in online contexts
- Compression and scalability (e.g.: user session reduction, spatio-temporal quantization, etc...)
- User data post-processing and manipulation by common image-processing algorithms
- Support for automatic locomotion models through *Ignited Location Signatures*

The next section will highlight related work, then section III will present and formalize session encoding techniques and layouts, including single and multiple user sessions. Section IV will discuss results of such techniques applied to different case studies, including public exhibits through desktop-based VR installation and online WebVR sessions, analyzing recording and encoding processes, with local or remote visual inspection.

II. Related Work

Nowadays, several VR head-mounted displays (HMDs) available at consumer level can offer incredible visual clarity and sense of presence within simulated 3D scenes. Oculus

978-1-7281-0292-4/18/$31.00 ©2018 IEEE

Rift, Sony PlayStation VR, HTC Vive and others are just a few examples of such headsets, often used during public exhibits, allow users to explore virtual environments (VEs). Within the Cultural Heritage field, HMDs employed in public spaces (museums, exhibits, etc.) can offer users really immersive experiences applied to virtual reconstructions, although limited by several constraints (short timings to name one). Capturing, storing, mining and visualizing the whole history of all user sessions can offer valuable insights about the experience for a specific 3D environment. Management of massive amounts of recorded data related to users sessions [1] within museums or public exhibits presents several challenges - especially if we deal with immersive VR and data transport in online contexts. In previous work data mining methods have been mostly applied to record tourist activity for discovering landmark preferences from photo mappings [2] where classical clustering methods [3] can also be employed to analyse spatial behaviors. Analyzing serialized explorations for thousands of casual visitors interacting with different 3D virtual environments through an HMD can offer valuable insights to understand user attitudes, spatial affordances, attention and focus within well-defined spatial extents. Recent research trends also investigate immersive information visualization [4] to support analytical reasoning on multi-dimensional data. Furthermore, recent works also investigate impacts of physical workload applied to use of HMD, including neck fatigue [5], rehabilitative applications using altered visual feedback to influence movement [6] and studies on physical strain for prolonged use of HMD [10] by measuring muscle activity of neck and shoulders. Furthermore, locomotion applied to consumer VR devices is actively investigated by plenty of studies as it is crucial to create an experience that keeps users comfortable during the session. Most VR experiences through HMDs often put limitations on the ability to walk or use specific locomotion techniques [9] to avoid well-known effects such as simulator sickness [7]. Recent literature reviews [8] also demonstrated concrete connections between locomotion and sense of presence, thus leading to large impacts for immersive experiences. Collected raw data and its analysis can thus play a big role to understand and optimize the VR exploration itself. For instance such data can offer support to shorten session duration by applying scene-dependent locomotion constraints in valuable locations, or even shifting the whole locomotion technique for a given 3D scene. Recent developments within online immersive fruition, fueled the expansion of the open specification WebVR (soon WebXR) making it possible to experience immersive VR using a common browser. The main goal is to make it easier for everyone to get into immersive VR experiences, abstracting from hardware and the wide range of available HMD devices on consumer market. Well-established commercial products like SketchFab [11] already employs such technologies to craft a platform to publish and present VR and AR content. A clear gap is still present with respect to desktop VR counterpart: performance, scene complexity and data streaming just to name a few. Visualizing or analyzing massive recorded data in a WebVR online context (e.g. user VR sessions being recorded in a remote location) presents several bottlenecks. The whole history of a single user session may contain huge amount of information (locomotion, orientation, etc) that may seriously prevent online transport due to data bandwidth. Regarding visualization of spatial marks within given 3D extents, plenty of past works already investigated volumetric data representation and visualization. Challenges often arise from trade-offs related to storage and performance, for instance the use of acceleration techniques to reduce per-fragment computations [12] or using out-of-core algorithms to build sparse voxel octrees from large input data [13]. Recent works also investigate solutions targeting modern GPUs and WebGL online contexts for mobile devices, for instance extending 2D mosaic approaches to store 3D data [14] and volume rendering on local client [15].

III. Session Encoding

This section defines and describes the proposed image-based encodings and lightweight layouts for VR sessions. The section discusses the advantages in terms of portability, storage, data quantization and transport also targeting modern web browsers and mobile.

A. User Session

Each user performs different interactions in a given time period, after activating the HMD to explore the Virtual Environment and before leaving (deactivating) the HMD. We formally define the user session operator *S* in order to query the state *s* of a given user u at given (global) time *t*:

$$S_u(t) \rightarrow s \tag{1}$$

The state s contains for instance the user HMD world location or orientation within the virtual scene, target, focal point or more sophisticated data. The state s can be also undefined if the user u was not present at time t (e.g. he/she left the VE). Within HMD sessions, we are generally interested in specific features like HMD location (P), HMD orientation (O), User focus (F), User scale (S) and derived attributes. Recording specific user states may provide insights on their attitudes at microscopic level, although at macroscopic level our objective is to measure the quality of *s* over time. In large or complex virtual scenarios with unconstrained locomotion techniques in fact, content rarefaction is very prone to occur, resulting in user disorientation and large efforts for a 3D modeling team.

B. Quantized User Session Volume (QUSV)

The general idea of the *Quantized User Session Volume* (QUSV) is to observe a portion of the virtual space and produce ”musical scores” where each ”note” encodes a specific user state. We define the QUSV as an AABB[1] structure defined by a unique ID, origin (world position) and 3D extents (width, length and height). The role of a single QUSV is to track and record specific user states within its volumetric extents. Such class is enriched with an *encoder*, able to write observed states

[1] Axis-Aligned Bounding Box

into compact image-based layouts (i.e.: our ”musical scores”). Regarding the observation of user spatial parameters, including for instance user location, focus, target, etc. the volume is subdivided into 3D cells (or voxels), thus each spatial feature is quantized depending on QUSV extents. Our approach is to employ a quantized encoding in order to map each 3D cell in the QUSV to a RGB value where R=x, G=y and B=z. Using a standard 8-bit per channel storage, the volume is subdivided into 256^3 voxels, thus each 3D location within the given QUSV can be color-coded (see Fig. 1).

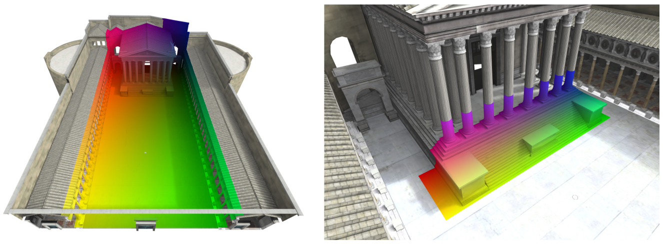

Fig. 1. two different QUSV. The left one has extents (57x120x70 m) thus each 3D cell has a size of 222 x 468 x 273 cm to record and keep track of main forum locations. The right QUSV is observing a smaller volume, the temple stairs (40x10x10 m) thus each 3D cell has size 150cm x 3cm x 3cm

The extents and scale of a QUSV have obviously a huge impact on single voxels size in world coords and thus on location quantization. The flexibility of such volumes is in fact to observe localized spatial extents with the purpose of capturing and encoding specific user behaviors.

C. Single user session

We can now define the encoding of $S_u(t)$ spatial location for a specific user *u* within a given QUSV as a lightweight mono-dimensional stream of RGB(A) values, as discrete signal - called QSS (Quantized Session Signal). For instance we can track and encode user HMD location as a variation of RGB values over time (x-axis) - see Fig. 2.

Fig. 2. sample QSS encoding a single user location over time as a signal (variation of RGB values). Image has been vertically stretched for clarity

With this approach, given a fixed temporal step increment, each texel of the signal encodes a specific location within the QUSV boundaries at given time t. For instance, if location recording was performed using 0.1s step, 10 contiguous texels of the QSS represent the user position variation within a single second. We can employ common 2D image processing algorithms to perform further data manipulation (see Fig. 3) or analysis, such as:

- isolate a single channel to analyse variation (e.g. blue channel for height variation)
- shrink temporal data using advanced resampling methods (e.g. bicubic resampling, etc.)
- apply 1-dimensional blur filtering to smooth spatial data
- employ lossy image compression algorithms

Fig. 3. sample 2D filters applied to user locomotion data: 1-dimensional blur (top) and isolated user HMD height variation (bottom)

Furthermore, we obtain a data layout suitable for encoding/decoding and manipulation directly on modern GPUs. For instance we are able to exploit available OpenGL ES texture filtering for automatic interpolation of spatial data at a given *t* via common texel fetching routines.

D. Multiple user sessions

At this point, we can stack multiple signals (QSS) into a 2D atlas, called QSA (Quantized Session Atlas): the QUSV encoder will take care of encoding and writing the entire $S_u(t)$ for a given feature. We propose a 2D layout where x-axis represents time and y-axis represent user ID (u). Such layout still allows offline image-processing operations and interactive manipulation on modern GPUs.

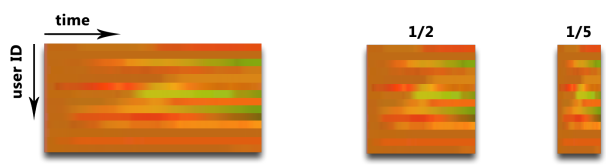

Fig. 4. sample QSA encoding locomotion data for 14 users and temporal compressions of 1/2 and 1/5 respectively. Session atlas has been vertically stretched for clarity

From a storage perspective, a QSA can be shrinked along the x-axis (temporal compression) using nearest-neighbor algorithm, thus allowing a coherent data reduction, where needed. Such reduction can be employed for different reasons, for instance to create $S_u(t)$ LODs, where progressive streaming can be used for massive datasets. From a real-time visualization perspective, such layout can be employed and easily manipulated by GPU as texture data, to transport and represent user or HMD states within the QUSV, including modern smartphones and tablets.

E. Temporal Patches

The size of the stream along x-axis (time) for QSS or QSA can be a limitation with respect to session duration: for instance imposing a time step of 0.1s and an available texture size of 4096, the maximum user session duration we are able to record is about 6 minutes. It is also true that for public exhibits users have generally limited timing for VR sessions due to serialization of the experience, so - under certain circumstances - this can be sufficient. For longer session durations, we can use temporal patches: once texture space is fully consumed, a new patch can be created and continue the recording.

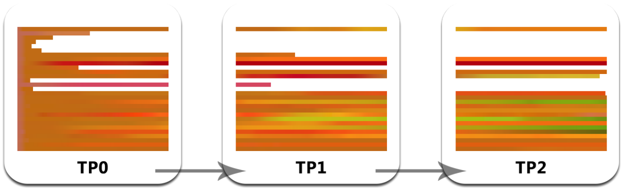

Fig. 5. temporal patches for prolonged user sessions

The recording process is thus able to allocate new patches and write them on disk, while a separate application can request a specific patch TP*k* depending on simulated time. Due to the lightweight layout, such segmentation is specifically suitable for both prolonged recording contexts and large amount of daily users. Furthermore, such compactness also enables online scenarios where the collector encodes running user sessions during the exhibit, while remote clients (e.g. web applications) request temporal patches. The latter approach can be exploited by professionals for direct visual inspection or third party data mining applications while the VR installation is running in the museum or expo.

F. Location Ignition and Signature

Spatial features within the QUSV (user location, target, focus, etc..) recorded over time can be exploited to *ignite* specific voxels of the virtual environment. Depending on hits, distance, persistence over time and other contributing factors, different locations or objects may present a variable rank/interest value for users. Large amount of data during exhibit can provide valuable ignition, and more interestingly, can offer a good estimate of real user propensities. Such data could be somehow synthesized to improve and optimize the VR session itself, for instance adjusting locomotion or fine-tuning interaction model. The general idea applied to spatial features (e.g. HMD world location, user focus, etc...) is to keep track of the number of hits for each voxel inside a specific QUSV and to maintain them sorted in a descending order.

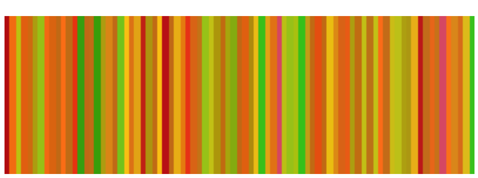

Fig. 6. a sample locomotion (HMD location) signature for all recorded users in a given VE. Leftmost values are most ignited voxels of the QUSV. Image stretched vertically for clarity

The algorithm is based on [16] and an hash-table to compute frequency counts (location ignition) while maintaining a constant memory footprint (k, the size of table) and a running record approach. In order to produce an "*ignited location signature*" (ILS) for a given QUSV, data is sorted by rank: for instance by using user focus, the leftmost value represents the most ignited voxel (computed from persistence over time and distance). Such sorted layout allows partial or progressive evaluation of the signature by discarding rightmost values (lower rank). For instance with k=512, we can partially evaluate the first half range of the signature (128 values) or just the single leftmost value (most ranked location). A single signature can even be generated from a single session, although valuable and useful data are clearly produced by multiple sessions.

IV. Experiment results and discussion

This section discusses different experiments, including visual and quantitative results where the presented image-based encoding techniques were applied, using an Oculus Rift CV1. For the first case, an open-source and portable VR explorer (ovrWalker [17]) was employed to collect user session data in standard CSV[2] files: such data was successively encoded through different QUSVs and compared both in terms of accuracy and storage. For the second case (WebVR) dedicated components (client and server-side) were developed to implement QUSV module and direct encoders for QSA and ILS. Regarding spatial ignition signatures we used k = 1024 (table size) and non-lossy PNG, consuming around 3 Kb of storage per signature.

A. First case: desktop-based VR application

Regarding the first case, the VR exploration was carried out through free linear locomotion techniques using a common joypad. Such locomotion methods may lead to motion sickness in some users due to the well-known conflict between visual and vestibular system [8], although it was part of the investigation. A few virtual environments were selected, already used in past exhibits [18] [19], each including collider geometries to support Physics (surface friction, gravity, etc...) for the locomotion method. A session recording component was developed to anonymously observe and record per-user session data during public exhibits for later analysis. The basic user state collected at given time intervals was composed by:

- *Timestamp* including also year, month and day of the record
- *HMD position* in world coordinates, including translation matrix from tracking sensors
- *HMD orientation* as quaternion

The state is silently written into a cumulative ASCII file via non-blocking callbacks in order to absolutely avoid unintended latency or skip HMD frames. Each user session handle is initiated by the component when user puts on the HMD and closed when its back into standby mode. The user session

[2]comma-separated values, ASCII file

collector was activated during "TourismA" 2018 event [20]: a 3-day international exhibition held in Florence (Italy) targeting all the cultural realities active in archaeological, artistic and monumental fields. The component recorded around 300 user sessions with time-step of 0.1s for each virtual environment. CSV files were produced on local machine disk for further data manipulation and analysis by third-party software. Collected records were finally encoded into presented image-based layouts (QSA, ILS) in order to:

- Perform storage comparisons (QSA and standard CSV) with large amount of casual users
- Test and compare spatial accuracy using different QUSV shapes and sizes
- Decode and visualize several user sessions on interactive WebGL Front-Ends at runtime
- Generate ignited location signatures for HMD location for each QUSV in each virtual environment

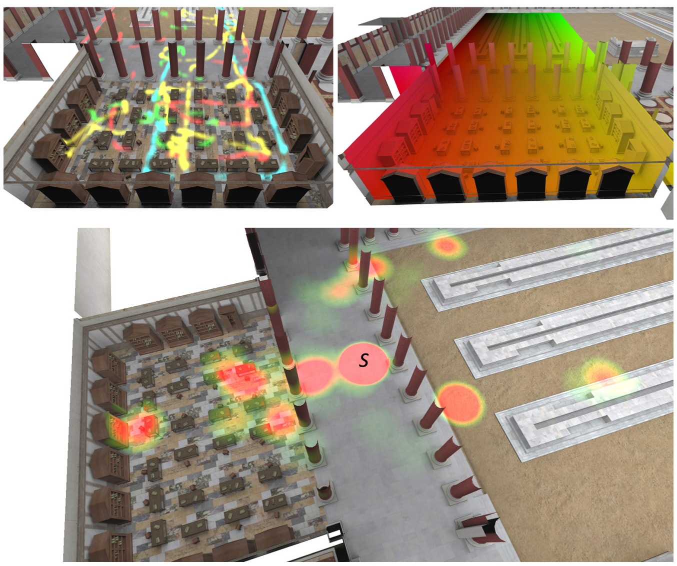

Fig. 7. 3D visualization of raw locomotion data from multiple CSV files for the "Forum Pacis" scene (top left); the resulting QUSV (color-coded for clarity) covering all walkable area (top right) and visualization of ignited location signature (HMD location) after 30 user sessions. S denotes the starting location for the experience

From multiple ILS evaluated with low ranges (16, 32 and 64), generated by all collected sessions, it was possible to automatically synthesise locomotion graphs for each scene, as direct result of users spatial propensities (location persistence over time). These graphs did allow to offer a compact, simplified locomotion model (node-based locomotion) over valuable areas for each virtual environment, thus removing collider geometries used with the old free locomotion model.

B. Second case: WebVR application

Regarding the second case (web-based scenario) a QUSV module has been implemented in a WebVR Front-End (javascript) provided by the open-source project ATON [21] [22] [23] [24] based on OSGjs library [25], the same used by

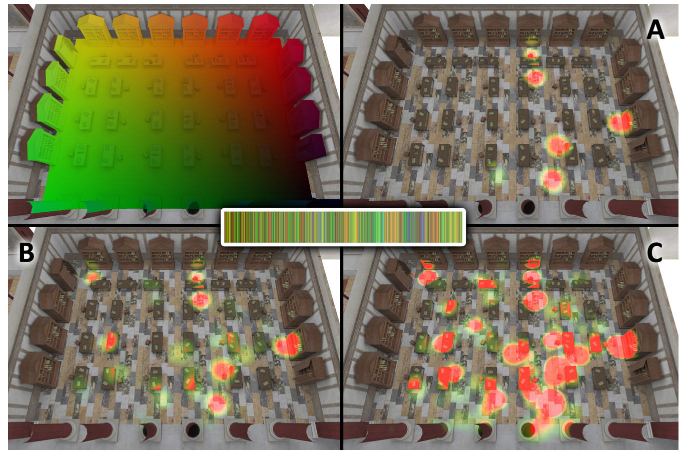

Fig. 8. A second QUSV to record finer locomotion data inside the library (top left, color-coded for clarity) and real-time GPU decoding of ignited location signature (HMD location) evaluated at different ranges (A=8, B=128 and C=512).

SketchFab platform. A corresponding server-side service (collector) based on node.js [26] was also developed to receive and process user states from connected clients through HTML5 web-sockets, recording and directly encoding incoming data following the presented techniques. Each state - sent via websocket message by the client - was composed by:

- *ID* (User ID)
- *HMD position* in world coordinates, including translation matrix from tracking sensors
- *HMD orientation* as quaternion
- *Focal point* in world coordinates

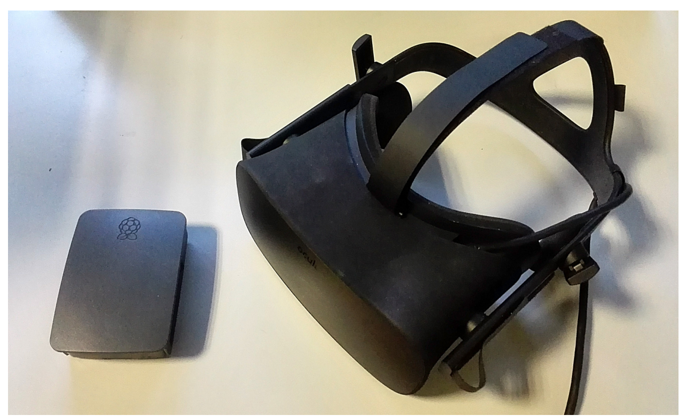

Fig. 9. The Raspberry Pi 3 (server) with running services for 3D content streaming, recording and encoding (left), and Oculus Rift CV1 (right)

The focal point is decoupled from actual HMD orientation, in order to support HMDs equipped with eye-tracking devices (e.g. Tobii [27]) for more accurate user focus. The resulting 3D point in world coords (including surface normal) is computed from the first intersection on the ray starting from HMD position and focus direction vector, thus automatically taking care of occlusion: for the experiment, we approximate such focus vector with HMD direction. Both the collector module (based on node.js) and the Front-End were deployed on a Raspberry Pi 3 board ($40) [28] [29], while WebVR experience was consumed by connected clients via local Wi-Fi hotspot from PC equipped with NVIDIA 980 GTX and i7 CPU, Oculus Rift CV1 and Firefox Quantum v60. While the current user was exploring the virtual environment, a professional was silently inspecting the same 3D scene with running QSAs and ignited signatures (HMD location and focus) served by the same Raspberry Pi. Regarding user interaction, the 3D web application this time offered users a teleport-based locomotion technique using HMD direction and common VR controllers. The actual location where the user is able to teleport is based on the intersection between HMD direction and an eligible surface. For the experiment we define a surface area eligible for teleport if its normal z-component satisfies the condition >0.8. The same Front-End was also used to load and visualize image-based encoded data through GLSL shaders and provide interactive web UI to control time, filter users or specific temporal intervals (see Fig. 10). The web component also allowed to validate the presented techniques and have a visual correspondence while inspecting data. Furthermore, collected datasets, including user sessions and ILSs could be also visualized through HMDs remotely. The ILS in this case was used to encode user focus (depending on distance and persistence) rather than HMD locomotion as in the desktop-based application. Distance contribution (HMD locaton to focal point) was computed by linearly increasing voxel ignition value under 5 meters, in order to give more impact to closer inspection. The experiment was carried out by 30 volunteers inside the Department of Computer Science (Sapienza University) on two virtual rooms, freely available (CC BY-SA 4.0) on SketchFab by The Hallwyl Museum [30].

Fig. 10. Web-based interactive visual inspection of locomotion data for two different users (red and yellow) decoded from a QSA. Notice the discontinuity between nodes due to teleport technique and local variations from positional tracking. A web user interface was also implemented to interactively filter locomotion history by user or session time using a slider.

For the first room, a single QUSV (12m x 14m x 7.3m) was used to observe running user sessions. Both storage and accuracy comparisons were performed between raw CSV data and encoded data. From a storage perspective, a first run with

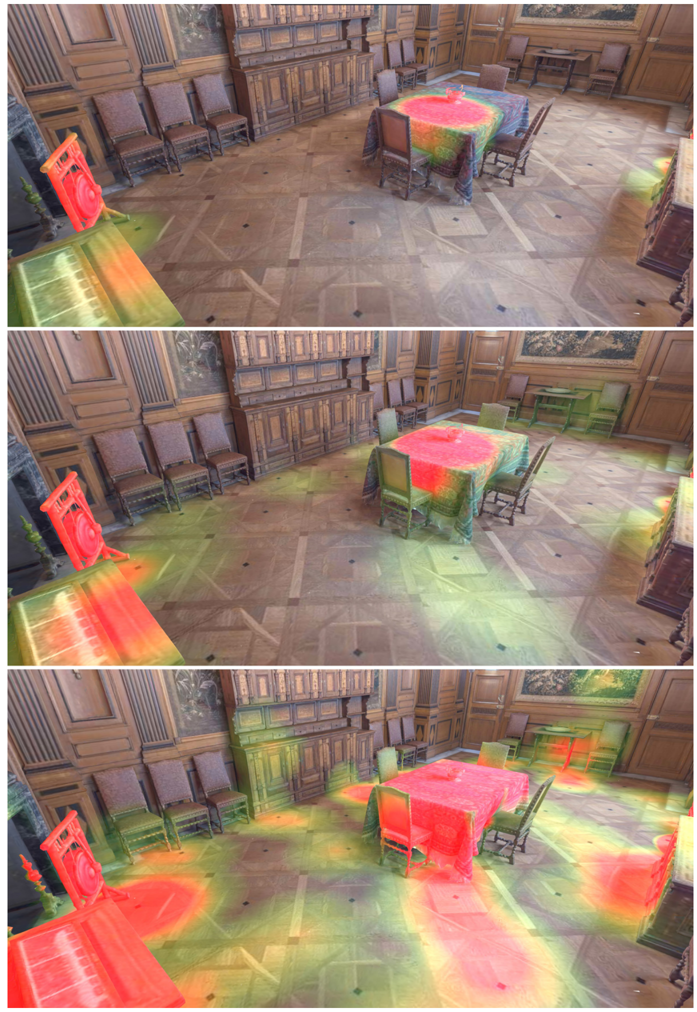

Fig. 11. the ILS decoded at runtime on the GPU after 5 WebVR sessions performed on the first sample room. ILS evaluated at ranges 16, 64 and 256 (from top to bottom)

2 users led to ~140 Kb of recorded CSV data, while image-encoded QSA was 4 Kb using non-lossy PNG format. Another run with 30 recorded users led to ~6.5 Mb while corresponding QSA around 10 Kb. From an accuracy perspective, the QUSV led to an error for spatial locations under 4 cm, 5 cm and 2 cm along X, Y and Z axes, respectively. For user locomotion analysis, such errors are quite tolerable, while regarding users focus it depends on objects or areas extents (e.g. a fork on the table). In the latter case, a second QUSV can be easily employed to observe a finer-grained volume with lower error, or even defined to fit a specific object or set of objects.

For the second room (see Fig. 12), a single QUSV (13m x 10m x 11.8m) was deployed to observe and encode multiple QSAs. Besides focus ignition thorugh computed signature, we employed multiple QSAs to understand additional behaviors: for this experiment HMD location, focus and HMD orientation were encoded (see Fig. 13). Notice how HMD location QSA is ”blocky” compared to focus QSA: this is related to locomotion technique used (teleport) thus resulting in a ”spatially

Fig. 12. the ILS decoded at runtime with range = 64 after 6 WebVR sessions on the second room. Notice how stairs are ignited because of the locomotion method (HMD gaze direction was used to teleport)

quantized" exploration where users perform small transitions and observe surroundings by local head motions. Specifically the latter is validated by higher frequencies in both focus and headset orientation QSAs over time: for head orientation a normalized direction was used, thus discarding roll motions (this can be captured by quaternion instead using RGBA).

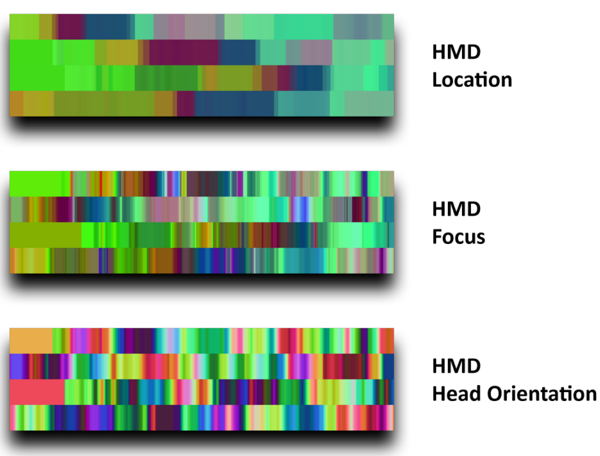

Fig. 13. Multiple QSAs encoded by the QUSV during 4 sessions on the second room: HMD location, focus and orientation.)

A more selective investigation was carried our on head motions, by employing comfortable motion zones discussed in literature on HMD ergonomics [31] thus encoding an additional QSA for head pitch, using following color codes:

- Very comfortable (-10°to 20°) - Blue
- Good (-40°to 60°) - Green
- Not comfortable (below -40°or above 60°) - Red

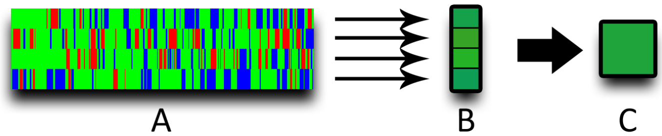

Fig. 14. comfort levels for head movements (range of motion zones) recorded in a QSA for 4 users (A), weighted average (B) and global average (C).

Fig. 14 shows resulting QSA (A) encoded in the same session in Fig. 13. Notice how sporadic "uncomfortable" head motions are due to fine details on rooves capturing users attention, while the locomotion technique (head direction used to point and teleport) also impacted on the final gradient. The central array (B) can be produced by simply applying a common 2D resampling filter compressing the image along x-axis (time) to obtain weighted average of head motion comfort for each user session, or a global average for all sessions (C). These values can be once again processed by offline image processing software or interactively read on the GPU for real-time visual inspection.

V. Conclusions and Future Work

Recording immersive VR sessions and inspecting visual data can offer valuable insights from several perspectives, especially when dealing with public installations (temporary or permanent) or online WebVR virtual scenes, to understand users propensities, spatial locomotion, focus, physical strain and much more. Within this work we proposed image-based techniques to record large amount of sessions into light, compact formats, using quantized approaches through flexible volumetric observers (QUSV) encoding session atlases similar to "musical scores" (QSA) and ignited location signatures (ILS). We discussed advantages in terms of storage and quantization targeting web transmission and streaming using temporal patches for prolonged sessions. We discussed applications of such data layouts to common offline image-based manipulation and real-time GPU encoding/decoding for information visualization and immersive analytics. Image-based layout also allows several advantages by leveraging on existing and robust compression algorithms. Described techniques can be easily replicated and applied to other 3D libraries and devices, ranging from desktop VR, online WebVR to mobile browsers. Recording and encoding techniques were applied and validated through a public event (first case) on casual visitors using a desktop-based VR experience, and through WebVR Front-End on sample scenes, also testing and investigating the efficiency of server-side encoding on low-cost boards (Raspberry Pi 3). Head orientation states as normalized direction (RGB) or quaternion (RGBA) provides also valuable information on head motions: for instance we can measure uncomfortable neck positions as in the second WebVR experiment thanks to well-known ranges in HMD literature, or measure head motion variance for a given session. Regarding ILS, well also improve the routine for automatic synthesis of locomotion graphs, by correlating ignited locations and focus, thus improving the quality of node selection. Furthermore, we'll explore image-based comparisons between multiple ILS - thus performed on 2D domain - to determine the level of session similarity between different users. Regarding focus specifically, an interesting direction will be to apply such techniques to HMDs equipped with eye tracking, like Tobii [27] or similar devices, and provide more accurate encoding of gaze behaviours in online WebVR contexts. The QSA encoding will be also particularly useful for future projects to record and inspect collaborative (synchronous) WebVR sessions due to its temporal layout and direct comparison performed on QSA y-axis (users), and how collaborative approaches may affect users spatial behaviours and QSAs. The results of this work will also directly contribute the the development of the open-source ATON 2.0 project, by offering modular components to record online WebVR sessions.

ACKNOWLEDGMENT

Authors want to thank the Hallwyl Museum and their 3D models freely available (CC BY-SA 4.0) on SketchFab; VisionLab at Computer Science department (Sapienza University) and testing students; D. Ferdani and E. Demetrescu (CNR ITABC) for samples of 3D models used for Keys2Rome project (http://keys2rome.eu/); A. Palombini, I. Cerato and L. Rescic (CNR ITABC) for ovrWalker installation support during "TourismA 2017" event.

REFERENCES

[1] Agus, Marco, et al. "Interactive 3D exploration of a virtual sculpture collection: an analysis of user behavior in museum settings." Proceedings of the 13th Eurographics Workshop on Graphics and Cultural Heritage (2016).

[2] Jankowski, Piotr, et al. "Discovering landmark preferences and movement patterns from photo postings." Transactions in GIS 14.6 (2010): 833–852

[3] Jain, Anil K. "Data clustering: 50 years beyond K-means." Pattern recognition letters 31.8 (2010): 651-666.

[4] Wagner Filho, Jorge A., et al. "Immersive Visualization of Abstract Information: An Evaluation on Dimensionally-Reduced Data Scatterplots." Proceedings of the 25th IEEE Conference on Virtual Reality and 3D User Interfaces (March 2018). Vol. 2. No. 3. 2018.

[5] Chihara, Takanori, and Akihiko Seo. "Evaluation of physical workload affected by mass and center of mass of head-mounted display." Applied Ergonomics 68 (2018): 204-212

[6] Chen, Karen B., et al. "Influence of altered visual feedback on neck movement for a virtual reality rehabilitative system." Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Vol. 58. No. 1. Sage CA: Los Angeles, CA: SAGE Publications, 2014.

[7] Ibnez, Manuel Lpez, and Federico Peinado. "Walking in VR: Measuring Presence and Simulator Sickness in First-Person Virtual Reality Games.". 2016

[8] Boletsis, Costas. "The New Era of Virtual Reality Locomotion: A Systematic Literature Review of Techniques and a Proposed Typology." Multimodal Technologies and Interaction 1.4 (2017): 24

[9] Habgood, Jacob, et al. "Rapid, continuous movement between nodes as an accessible virtual reality locomotion technique." (2018).

[10] Wille, Matthias, et al. "Prolonged work with head mounted displays." Proceedings of the 2014 ACM International Symposium on Wearable Computers: Adjunct Program. ACM, 2014.

[11] SketchFab: https://sketchfab.com/

[12] Kruger, Jens, and Rdiger Westermann. "Acceleration techniques for GPU-based volume rendering." Proceedings of the 14th IEEE Visualization 2003 (VIS'03). IEEE Computer Society, 2003.

[13] Baert, Jeroen, Ares Lagae, and Ph Dutr. "OutofCore Construction of Sparse Voxel Octrees." Computer Graphics Forum. Vol. 33. No. 6. 2014.

[14] Noguera, Jos M., and Juan-Roberto Jimnez. "Visualization of very large 3D volumes on mobile devices and WebGL." (2012).

[15] Congote, John, et al. "Interactive visualization of volumetric data with webgl in real-time." Proceedings of the 16th International Conference on 3D Web Technology. ACM, 2011.

[16] Manku, Gurmeet Singh, and Rajeev Motwani. "Approximate frequency counts over data streams." VLDB'02: Proceedings of the 28th International Conference on Very Large Databases. 2002.

[17] ovrWalker: http://osiris.itabc.cnr.it/scenebaker/index.php/projects/osglab/ovrwalker/

[18] Antonaci, Alessandra, and Alfonsina Pagano. "Technology enhanced visit to museums. A case study: Keys to Rome." proceedings of INTED2015, Madrid, Spain (2015): 2-4.

[19] Fanini, Bruno, et al. "Engaging and shared gesture-based interaction for museums the case study of K2R international expo in Rome." Digital Heritage, 2015. Vol. 1. IEEE, 2015.

[20] TourismA event: http://www.tourisma.it/home-2/

[21] ATON project: http://osiris.itabc.cnr.it/scenebaker/index.php/projects/aton/

[22] Meghini, Carlo, et al. "ARIADNE: a research infrastructure for archaeology." Journal on Computing and Cultural Heritage (JOCCH) 10.3 (2017): 18.

[23] Gonizzi, S., Malatesta, S., Fanini, B., et al. "The Winckelmann300 Project: Dissemination of Culture with Virtual Reality at the Capitoline Museum in Rome". International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 42.2, XLII-2, 371-378, https://doi.org/10.5194/isprs-archives-XLII-2-371-2018, 2018.

[24] Antal, Adriana, et al. "A complete workflow from the data collection on the field to the deployment of a Virtual Museum: the case of Virtual Sarmizegetusa." (2016).

[25] OSG.js: http://osgjs.org/

[26] NodeJS: https://nodejs.org/en/

[27] Tobii VR: https://www.tobii.com/tech/products/vr/

[28] Raspberry Pi 3: https://www.raspberrypi.org/products/raspberry-pi-3-model-b/

[29] Vujovic, Vladimir, and Mirjana Maksimovic. "Raspberry Pi as a Wireless Sensor node: Performances and constraints." Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on. IEEE, 2014.

[30] https://sketchfab.com/TheHallwylMuseum

[31] Chu, Alex. "VR Design: Transitioning from a 2D to a 3D Design Paradigm." Samsung Developer Conference. Vol. 19. 2014.