



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA IN INGEGNERIA INFORMATICA

VALUTAZIONE SOGGETTIVA DELLA QUALITÀ DI POINT CLOUD COMPRESSI

Relatrice: Prof.ssa Federica Battisti

Laureando: Andrea Campagnol

Matricola: 1163671

ANNO ACCADEMICO 2021-2022

Data di laurea 22/07/2022

*A nonno Piero,
che mi sta accanto ogni giorno.*

Introduzione

Il progresso nelle tecnologie multimediali sta promuovendo l'adozione di media sempre più immersivi e fotorealistici, di facile accesso e trasmissione. La tecnologia Point Cloud si è rivelata uno strumento efficace e versatile per la rappresentazione volumetrica di oggetti e scene in ambienti tridimensionali, trovando applicazione in molti settori. Questa tecnologia è tuttavia ancora molto giovane e porta con sé diverse problematiche. I point cloud possono contenere milioni di punti. Nel loro formato originale, infatti, richiedono grandi quantità di memoria per l'archiviazione o larghezza di banda per la trasmissione. Per rendere utilizzabili i Point Cloud, la compressione è quindi necessaria. Questo processo comporta inevitabilmente la perdita di una certa quantità di informazioni, è utile quindi studiare l'effetto della compressione sulla qualità dell'esperienza percepita dall'utente finale. L'analisi statistica dei giudizi soggettivi forniti da questi utenti rappresenta il miglior punto di partenza per lo sviluppo e il miglioramento delle tecnologie coinvolte e permette di trovare il trade-off migliore tra gli obiettivi di funzionalità prefissati e i requisiti qualitativi richiesti. L'obiettivo di questo lavoro di tesi è raccogliere e analizzare dati soggettivi relativi alla QoE nella visualizzazione di Point Cloud compressi in ambiente VR, con particolare attenzione alla componente geometrica e strutturale dei modelli Point Cloud.

Indice

Introduzione	5
Listing of Acronyms	10
1 Tecnologie immersive	13
1.1 Tecnologie immersive	13
1.2 Supporti di visualizzazione	14
1.3 Rappresentazione tridimensionale	16
2 Point Cloud	19
2.1 Definizione di Point Cloud	19
2.2 Casi d'uso	19
2.3 Modalità di acquisizione	21
2.4 Modalità di rappresentazione e compressione	21
3 Metodologie di valutazione della Qualità dell'esperienza per i Point Cloud	25
3.1 Tecniche di valutazione oggettiva	25
3.1.1 Metrica geometrica Point-to-point PSNRD1	26
3.1.2 Metrica geometrica Point-to-plane PSNRD2	27
3.1.3 Similarità angolare Plane-to-plane	27

3.1.4	Metrica Point-to-point per gli attributi	28
3.2	Tecniche di valutazione soggettiva	29
3.2.1	Single Stimulus SS	30
3.2.2	Double Stimulus DS	30
3.3	Analisi dei dati	31
4	Esperimento soggettivo	33
4.1	Obiettivi	33
4.2	Partecipanti	33
4.3	Apparatus	34
4.4	Dataset	34
4.4.1	Selezione dei modelli	34
4.4.2	Decimazioni	35
4.4.3	Dataset finale	35
4.5	Esperimento	36
4.5.1	Protocollo	36
4.5.2	Ambiente virtuale	36
4.5.3	Esecuzione	37
4.5.4	Osservazioni	39
5	Analisi dei dati	45
5.1	Dati raccolti	45
5.2	Operazioni preliminari: outliers	46
5.3	Mean Opinion Score: MOS	46
5.3.1	MOS per Point Cloud	46
5.3.2	MOS per livello di dettaglio	51
5.4	T-Test	53

5.5 Dati secondari	53
Conclusioni	55
Bibliografia	57
Ringraziamenti	59

Listing of acronyms

ACR	Absolute Category Rating
ACR-H	Absolute Category Rating with Hidden reference
AR	Augmented Reality
DS	Double Stimulus
DSCQS	Double Stimulus Continuous Quality Scale
DSIS	Double Stimulus Impairment Scale
FPS	Frame Per Secondo
G-PCC	Geometry-based Point Cloud Compression
HMD	Head Mounted Display
JND	Just-Noticeable Difference
LOD	Level Of Detail
MOS	Mean Opinion Score
MPEG	Moving Picture Expert Group
MR	Mixed Reality
MSE	Mean Square Error
PC	Point Cloud
PCC	Point Cloud Compression
PSNR	Peak Signal to Noise Ratio
QoE	Quality of Experience
SDSCE	Simultaneous Double Stimulus for Continuous Evaluation
SS	Single Stimulus
SSCQE	Single Stimulus Continuous Quality Evaluation

Listing of Acronyms

V-PCC Video-based Point Cloud Compression

VR Virtual Reality

Capitolo 1

Tecnologie immersive

La multimedialità si è affermata come una delle più efficaci forme di rappresentazione e diffusione dell'informazione. Il progresso nelle tecnologie di acquisizione e visualizzazione multimediale sta promuovendo l'adozione di media sempre più immersivi e fotorealistici, di facile accesso e trasmissione. Queste caratteristiche aprono le porte a nuove opportunità di innovazione in molti settori ma portano con sé molte sfide. È necessario, infatti, garantire all'utente una fruizione dei contenuti efficiente, senza interruzioni o ritardi, ma allo stesso tempo piacevole, coinvolgente e senza degradazioni di qualità. Trovare un buon compromesso tra questi due aspetti è uno dei problemi principali delle applicazioni multimediali ed in particolare per le più recenti tecnologie immersive.

1.1 Tecnologie immersive

Con il termine “tecnologia immersiva” si fa riferimento all'insieme di tecnologie hardware e software che hanno lo scopo di fondere, o connettere in modo interattivo il mondo reale con quello virtuale. Le tecnologie immersive generano un'esperienza sensoriale simulata in cui l'utente può interagire con il mondo virtuale tramite delle apposite interfacce. Esistono diversi gradi di immersività, che è possibile ricondurre a 3 macrocategorie: Realtà Virtuale (VR), Realtà Aumentata (AR) e Realtà Mista (MR). La divisione tra queste classi non è netta, in quanto il livello di interazione e immersione in ogni categoria cambia molto in relazione alle caratteristiche dei sistemi e delle tecnologie utilizzati; tuttavia, è possibile delinearne gli aspetti caratteristici.

I sistemi di realtà aumentata combinano elementi reali e virtuali, tipicamente sovrapponendo o posizionando oggetti tridimensionali in ambienti e scene reali. La realtà viene quindi arricchita o “aumentata” con informazioni generate al computer.

Nei sistemi di realtà virtuale l'utente è completamente immerso in un ambiente digitale, che rimpiazza quello reale, tipicamente tramite l'uso di un visore e controller appositamente progettati per questo tipo di applicazioni. In questo caso la sensazione di immersività è totale e la percezione della realtà è completamente simulata.

La realtà mista è più difficile da definire poiché possiede caratteristiche sia della VR che dell'AR. In questi sistemi la realtà non viene rimpiazzata come nella VR ed il grado di interazione è più alto rispetto ai sistemi di AR. Gli oggetti virtuali sono infatti integrati nel mondo reale e possono interagire con esso e con l'utente come se fossero parte della realtà.

1.2 Supporti di visualizzazione

Esistono vari supporti per la fruizione di un'esperienza immersiva, le caratteristiche dei quali dipendono dal grado di immersione e interazione che si vuole ottenere.

Applicazioni di **Augmented Reality** tipicamente fanno uso di due elementi principali: telecamera e schermo. Le informazioni virtuali vengono inserite nell'ambiente inquadrata dalla telecamera e visualizzate a schermo. Algoritmi di tracciamento, intelligenza artificiale e sensori di movimento permettono di scalare e posizionare gli oggetti 3D nella scena in modo tale che sembrino realmente presenti. Dispositivi come gli smartphone sono ideali per eseguire questo tipo di applicazioni. Il grado di immersione è comunque limitato, infatti la realtà viene percepita come solo aumentata o arricchita dagli elementi virtuali e l'interazione con essi è possibile solo attraverso lo schermo, ad esempio tramite il tocco su un display touchscreen.

In un'esperienza di **realtà virtuale** il grado di immersione è decisamente maggiore e le tecnologie usate sono più sofisticate come i visori HMD (Head Mounted Display). Un visore per la VR è composto da uno o due schermi, preceduti da due lenti poste davanti agli occhi in grado di ricreare l'effetto di stereoscopia per illudere il cervello umano di percepire immagini tridimensionali. I visori HMD 3-GdL (a 3 gradi di libertà) montano dei sensori quali giroscopio e accelerometro per poter tracciare la rotazione della testa nelle 3 direzioni degli assi X,Y,Z. I visori più sofisticati 6-GdL sono in grado di tracciare anche il movimento nello spazio, così da permettere non solo la visualizzazione in tutte le direzioni della scena ma anche il movimento all'interno di essa spostandosi fisicamente. Per aumentare il grado di immersività spesso questi sistemi sono accompagnati da ulteriori tecnologie come cuffie per l'audio spaziale, o sistemi di tracciamento dell'occhio così da poter mettere a fuoco la porzione di immagine effettivamente osservata e ricreare un



Figura 1.1: AR application

effetto visivo più realistico. Altre tecnologie in grado di simulare un'esperienza di realtà virtuale sono ad esempio i monitor 3D e le stanze immersive. I monitor 3D, tuttavia, non forniscono un alto grado di immersività nonostante l'effetto di profondità che sono in grado di riprodurre e il campo visivo è fortemente limitato dalle dimensioni dello schermo. All'interno delle stanze immersive, invece, l'ambiente virtuale viene proiettato in scala 1:1 sulle pareti, immergendo l'utente nel mondo virtuale senza la necessità di visori HMD. Questo tipo di simulazione non è adatta alla diffusione di massa, viene infatti usata principalmente per esposizioni artistiche o museali.



Figura 1.2: VR headset and controller Meta Quest 2

Per la **realtà mista** vengono sfruttate entrambe le tecnologie di AR e VR. Un ottimo esempio è dato dal visore HMD Microsoft hololens 2. In questo visore non sono presenti schermi ma lenti olografiche trasparenti che “filtrano” la visuale dell'utente: gli oggetti

virtuali sono proiettati sulle lenti, interponendosi tra occhio e ambiente aumentando la realtà, in modo simile alla AR. Sistemi di tracciamento della mano, comandi vocali, tracciamento oculare e mappatura spaziale permettono di interagire con gli oggetti 3D senza l'uso di ulteriori controller o interfacce.



Figura 1.3: MR headset Hololens 2

1.3 Rappresentazione tridimensionale

Il Design e la funzionalità delle componenti hardware sono di fondamentale importanza per la resa di un'esperienza immersiva; tuttavia, è necessario fornire a tali sistemi dei modi per rappresentare realtà ed elementi virtuali in modo consistente all'esperienza che si vuole simulare. La rappresentazione tridimensionale si adatta bene ai requisiti di realismo, profondità e senso di presenza di cui necessitano le applicazioni immersive. Il rendering 3D ha lo scopo di approssimare le forme e le superfici continue di oggetti ed ambienti in modo discreto, e questo obiettivo può essere raggiunto utilizzando metodi di rappresentazione tridimensionale diversi.

- **Voxel:** Il solido è rappresentato tramite il suo volume, definito da delle unità volumetriche discrete dette voxel. L'oggetto tridimensionale è quindi il risultato di una suddivisione di quello originale in unità tridimensionali di base che ne compongono complessivamente il volume. Si può pensare ai voxel come elementi 3D che svolgono la stessa funzione dei pixel per un'immagine 2D.
- **Mesh:** il solido è rappresentato tramite la sua superficie, approssimata da dei punti collegati tra loro chiamati vertici. Le interconnessioni generano dei poligoni

adiacenti, tipicamente quadrilateri o triangoli, che nel complesso vanno a costituire la superficie dell'oggetto rappresentato. Più il numero dei vertici è alto meglio viene approssimata la superficie, a discapito della complessità computazionale necessaria per il rendering e l'occupazione di memoria.

- **Point Cloud:** il solido viene rappresentato da una nuvola di punti grezzi nello spazio che ne delineano forma e volume. I dettagli di questo tipo di rappresentazione sono esposti nel capitolo successivo in quanto oggetto di studio di questa tesi.

Capitolo 2

Point Cloud

2.1 Definizione di Point Cloud

Un Point cloud (PC) è una rappresentazione tridimensionale di un oggetto o di una scena definita da un insieme di punti in un determinato sistema di coordinate. Ogni punto è caratterizzato da una posizione nello spazio descritta da delle coordinate (e.g. secondo gli assi x , y , z) e da eventuali attributi o valori di intensità (e.g. RGB, luminanza, crominanza. . .). A differenza delle mesh poligonali, che contengono informazioni sulla connessione o sulla parentela tra vertici, i Point Cloud descrivono dati spaziali grezzi, non strutturati in quanto le relazioni spaziali tra punti non sono rappresentate. Le coordinate dei punti possono essere espresse da valori in virgola mobile o approssimati a valori interi. La quantizzazione delle coordinate, secondo la precisione spaziale richiesta, genera una griglia tridimensionale costituita da porzioni di volume dette Voxel. Tutti i punti quantizzati che risiedono all'interno di una stessa porzione di volume vengono mappati nelle coordinate del punto centrale del voxel. Il processo di conversione delle coordinate spaziali da virgola mobile a valori interi è chiamato voxelizzazione.

2.2 Casi d'uso

La tecnologia Point Cloud (PC) si è rivelata uno strumento efficace e versatile per la rappresentazione volumetrica di oggetti e scene in ambienti tridimensionali, trovando applicazione in molti settori [1]:

- Virtual, augmented e mixed reality
 - Computer grafica e gaming

2. Point Cloud

- Telepresenza
- Creazione di contenuti 3D
 - Design, progettazione e 3D printing
- Applicazioni mediche
 - Imaging medico 3D
 - Progettazione e produzione di protesi e parti del corpo
- Costruzioni e settore manifatturiero
 - Analisi della struttura 3D per il rilevamento dei difetti
 - Scansione 3D per supportare la gestione dei progetti
 - Riparazione e analisi di luoghi confinati o pericolosi
 - Rilevamento di crepe e danni per strade, ponti e altre infrastrutture
 - Pianificazione urbana e analisi urbana
- Vendita al dettaglio e beni di largo consumo
 - Modellazione di piccoli oggetti: gioielli, decorazioni, scarpe, ecc.
 - Modellazione di oggetti di medie dimensioni (auto, moto, mobili, ecc.) e oggetti di grandi dimensioni (case, appartamenti, ecc.)
- Conservazione del patrimonio culturale:
 - Scansione e modellazione di dipinti, gioielli, ceramiche, ossa, fossili e altri manufatti di piccole e grandi dimensioni
- Telerilevamento e sistemi informativi geografici
 - Scansione terrestre o aerea ad ampio raggio
- Sistemi di navigazione automatica:
 - Veicoli autonomi
 - Robotica
- Sorveglianza

2.3 Modalità di acquisizione

I Point Cloud possono essere ottenuti in molteplici modi e da diverse sorgenti. Una possibile classificazione distingue metodologie di acquisizione dirette (o attive) e indirette (o passive) [2].

- **Acquisizione diretta:** appartengono a questa categoria sistemi di imaging specifici per la raccolta di informazioni tridimensionali, sensori appositamente progettati per la scansione e la collezione di dati tridimensionali. I metodi attivi utilizzano sorgenti luminose (infrarossi o laser) e luci riflesse retrodiffuse per misurare le distanze tra gli oggetti e il sensore. I valori di profondità misurati vengono associati ai punti che compongono la scena generando un point cloud. Questa categoria include i sensori LiDAR, Radar e Telecamere a tempo di volo. Esempi di sensori di profondità attivi sono Microsoft Kinect, Apple TrueDepth Camera, Intel Realsense, Sony DepthSense e molti altri.
- **Acquisizione indiretta:** appartengono a questa categoria metodi che si appoggiano ad algoritmi di derivazione di informazioni spaziali da modalità di imaging che non acquisiscono direttamente informazioni 3D. La generazione di PC da fonti non tridimensionali, come un set di immagini 2D standard, si basa sulla ricerca di corrispondenze tra punti in immagini diverse dello stesso soggetto (o scena) scattate da più angolazioni e sul calcolo di triangolazioni spaziali per dedurre la distanza tra gli oggetti nello spazio 3D. I point cloud possono inoltre essere ricavati artificialmente da modelli 3D digitali, riducendo ad esempio una mesh poligonale ai soli vertici. Esempi di metodologie caratteristiche di questa classe sono la fotogrammetria con più telecamere, i light field e le derivazione tridimensionali “Structure from motion” e “shape from shading”.

I metodi di acquisizione attiva e passiva possono essere utilizzati in modo complementare per migliorare la generazione delle nuvole di punti e raggiungere una precisione maggiore.

2.4 Modalità di rappresentazione e compressione

Le applicazioni in cui i Point Cloud sono utilizzati direttamente da esseri umani richiedono il rendering per poter essere visualizzati, procedura che può basarsi su diversi modi di rappresentazione finale di un Point Cloud.

- **Rappresentazione tramite punti:** è la più semplice modalità di rappresentazione, il point cloud viene visualizzato solo tramite i suoi punti grezzi, nessuna superficie viene approssimata. Gli attributi, come il colore, possono essere visualizzati singolarmente in ogni punto. Questo tipo di visualizzazione è adatto per applicazioni con requisiti di memoria stringenti e per point cloud molto densi.

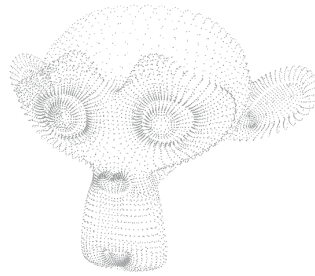


Figura 2.1: Point representation

- **Rappresentazione implicita:** i punti del point cloud vengono approssimati da superfici locali $f(x,y,z)$, ognuna delle quali è definita in termini delle sue coordinate cartesiane x , y , z . Queste superfici possono essere visualizzate direttamente utilizzando metodi di ray tracing o dopo la conversione in una rappresentazione basata su poligoni e proiettate sul piano di osservazione 2D. Il calcolo delle superfici è complesso e richiede informazioni sui vettori normali per ogni punto del PC. Se queste informazioni non sono disponibili occorre stimarle rendendo il metodo ancora più complesso.

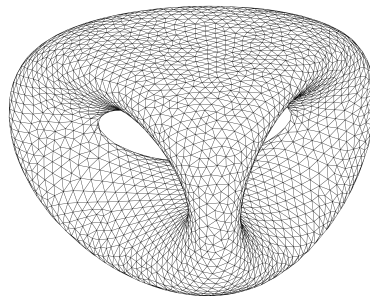


Figura 2.2: Implicit representation: Surface of genus 2: $2y(y^2 - 3x^2)(1 - z^2) + (x^2 + y^2)^2 - (9z^2 - 1)(1 - z^2) = 0$

- **Rappresentazione esplicita:** La superficie del point cloud viene approssimata da semplici poligoni, ottenuti ad esempio dalla triangolazione di Delauney o dalla

decomposizione delle regioni di Voronoi. In generale, i punti del Point Cloud vengono connessi tra loro, fungendo da vertici per i poligoni che si formano. Questi, adiacenti tra loro, approssimano nel complesso una superficie continua.

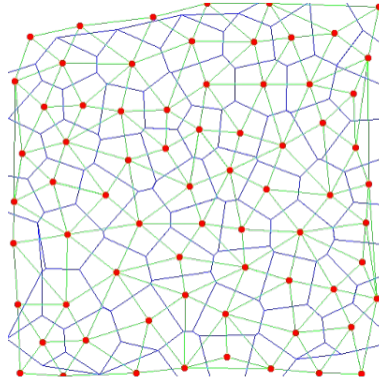


Figura 2.3: Explicit representation

- **Rappresentazione tramite Punti di dimensione variabile o splats (macchie):** Se la nuvola di punti è molto densa, invece di calcolare una superficie continua approssimata, il rendering può essere ottenuto utilizzando quadrati o cubi di dimensioni variabili o piccole superfici ellissoidali associate ad ogni punto del PC per ottenere una superficie in apparenza continua e priva di buchi, che può essere dipinta con il colore dei punti PC.



Figura 2.4: Splats representation

I point cloud possono contenere milioni di punti, per ognuno dei quali vengono memorizzate coordinate ed eventuali attributi. Nel loro formato originale, infatti, i Point Cloud

richiedono grandi quantità di memoria per l'archiviazione o larghezza di banda per la trasmissione. Nel caso di Point Cloud dinamici il tutto si traduce in milioni di punti per ogni frame; i problemi dell'occupazione di memoria e la capacità di trasmissione diventano quindi rilevanti. Lo sviluppo di tecnologie di acquisizione di Point Cloud a più alta risoluzione, inoltre, impone un requisito sempre più elevato sulla loro dimensione. Per rendere utilizzabili i Point Cloud, la compressione è quindi necessaria. Il Moving Picture Expert Group (MPEG) ha selezionato due distinte tecnologie di compressione (PCC) come standard di compressione per i Point Cloud: video-based Point cloud compression (V-PCC) per i point cloud dinamici e geometry-based Point cloud compression (G-PCC) per quelli statici. I metodi V-PCC convertono i dati del Point Cloud proiettandoli dallo spazio 3D in piani 2D, che vengono poi codificati con strumenti tradizionali per la codifica di video 2D. Gli algoritmi G-PCC si basano sulla decomposizione octree per codificare la geometria del point cloud direttamente nello spazio 3D.

L'uso di algoritmi di compressione può introdurre artefatti o distorsioni ed impattare la qualità percepita dall'utente durante la visualizzazione. Inoltre, altre degradazioni possono apparire durante il processo di acquisizione, rendering e trasmissione.

Acquisizione: potrebbero verificarsi errori nella posizione dei punti e/o perdite di punti.

Codifica: Con l'applicazione dell'algoritmo G-PCC potrebbero essere creati falsi bordi o essere persi alcuni punti sull'intera figura. Con l'uso dell'algoritmo V-PCC, gli artefatti più visibili sono la blocchetizzazione e i falsi bordi. Inoltre nei codec Point Cloud, alla diminuzione del bitrate target (profondità octree inferiore), corrisponde una diminuzione nel numero di punti decodificati, poiché, tutti i punti all'interno di un voxel vengono rappresentati da un solo punto. La conseguenza è un aumento della distanza tra i punti decodificati e, quindi, un decremento nei dettagli.

Rendering: le distorsioni possono sorgere a seconda della dimensione dell'elemento primitivo utilizzato per la ricostruzione del PC: un elemento più piccolo potrebbe comportare la creazione di fori, mentre uno più grande potrebbe causare l'aliasing sul PC renderizzato. In altri termini, un PC decodificato può risultare pixelato o eccessivamente sottocampionato [3].

Trasmissione: Degradazioni visive nei point cloud possono sorgere in relazione alla perdita o a ritardi nella ricezione dei dati. I canali di trasmissione, inoltre, possono introdurre rumore o causare errori che possono alterare il contenuto dei file PC. In letteratura si trovano poche informazioni sugli effetti di queste problematiche.

Capitolo 3

Metodologie di valutazione della Qualità dell'esperienza per i Point Cloud

Come già discusso nel capitolo precedente, i processi di acquisizione, compressione, trasmissione e rendering possono produrre delle degradazioni visive dei point cloud che vanno ad impattare notevolmente sulla qualità dell'esperienza percepita dall'utente. È importante riuscire a dare una misura a questa percezione per migliorare o raffinare tecnologie e algoritmi in determinate condizioni d'uso e trovare il miglior compromesso tra qualità e funzionalità. Le tecniche di valutazione della qualità mirano a risolvere al meglio questo problema, combinando strumenti matematici e opinione umana. Si possono distinguere rispettivamente metriche oggettive e metriche soggettive. Idealmente le misure di qualità oggettive dovrebbero coincidere con i risultati dei test soggettivi, cioè con la percezione dell'utente.

3.1 Tecniche di valutazione oggettiva

Le tecniche di valutazione oggettive sono metriche che cercano di predire la qualità visiva calcolando la somiglianza tra il contenuto distorto in esame e quello di riferimento utilizzando una misura oggettiva, cioè calcolata matematicamente. Questi metodi di misurazione considerano le caratteristiche del sistema visivo nel tentativo di dare misure di qualità coerenti con il giudizio umano. Nel caso dei point cloud si possono valutare 3 aspetti principali che vanno ad incidere sulla qualità: artefatti geometrici, degradazioni degli attributi, e disturbi dovuti al bit rate. In questa trattazione le metriche descritte

fanno riferimento alla geometria e agli attributi dei PC, le componenti in cui gli artefatti possono causare una percezione negativa della qualità più elevata. La maggior parte di queste tecniche di valutazione si basa su metriche già utilizzate per immagini/video, come il Peak Signal to Noise Ratio (PSNR) e MSE [4].

$$PSNR = 10 \log \left(\frac{p^2}{MSE} \right) \quad (3.1)$$

$$MSE = \frac{1}{MN} \sum_{i=1}^N \sum_{j=1}^M [I(i, j) - I'(i, j)]^2 \quad (3.2)$$

Dove I è l'immagine originale e I' è l'immagine distorta. M e N sono la larghezza e l'altezza dell'immagine. p è la gamma dinamica dei valori dei pixel.

3.1.1 Metrica geometrica Point-to-point PSNRD1

La metrica PSNRD1 si basa sulla distanza geometrica tra i punti di un PC in fase di valutazione, B , e i punti associati del PC di riferimento, A . Nello specifico, per ogni punto b_j del point cloud degradato B , viene individuato il corrispondente punto a_j appartenente a quello originale A , attraverso l'algoritmo nearest neighbor. Le distanze tra i punti a_j e i corrispondenti punti b_i determinano il vettore errore $E(i, j)$ dove il singolo errore Punto-punto è calcolato con la distanza euclidea:

$$e_{B,A}^{D1}(i) = \|E(i, j)\|_2. \quad (3.3)$$

L'errore Point-to-point (D1) per l'intera nuvola di punti è quindi definito come la media aritmetica, $e_{B,A}^{D1}$, degli errori $e_{B,A}^{D1}(i)$

$$e_{B,A}^{D1} = \frac{1}{N_B} \sum_{i=1}^{N_b} e_{B,A}^{D1}(i) \quad (3.4)$$

Questo errore può essere anche espresso dalla distanza di Hausdorff: definita come la massima distanza tra tutti i punti associati b_i e a_j [5]:

$$d_{haus} = \max_{i \in B} e_{B,A}^{D1}(i) \quad (3.5)$$

L'errore Point-to-point $e_{B,A}^{D1}$ viene espresso come PSNRD1 simmetrico usando la seguente formula:

$$PSNRD1 = 10 \log_{10} \left(\frac{3 * peak^2}{\max(e_{B,A}^{D1}, e_{A,B}^{D1})} \right) \quad (3.6)$$

dove *peak* è la risoluzione geometrica del modello (cioè se voxel bit depth = 10, è 1024) e $e_{A,B}^{D1}$ è l'errore Point-to-point quando i ruoli di A e B sono scambiati in Equazione (3.3). Il denominatore dell'equazione (3.6) assicura che il PSNR risulti simmetrico ed invariante rispetto a quale dei due PC è considerato come riferimento.

3.1.2 Metrica geometrica Point-to-plane PSNRD2

La metrica PSNRD2 si basa sulla distanza geometrica tra i punti associati di un point cloud oggetto della valutazione, B, e i piani locali generati dai punti del point cloud di riferimento A. Come per la PSNRD1, per ogni punto b_i del Point Cloud B, viene identificato il suo corrispettivo a_j nel PC di riferimento A, attraverso l'algoritmo nearest neighbor, e calcolato il vettore errore $E(i, j)$. Per ogni punto a_j di A viene interpolato un piano per la regione che ha centro in tale punto e calcolato il vettore normale N_j . Questo viene fatto usando ad esempio un modello quadratico di fitting della superficie con un raggio di e.g. 20 punti vicini (parametro variabile a seconda della densità del PC). Il vettore errore $E(i, j)$ tra i punti b_i e a_j viene proiettato sul vettore normale associato N_j attraverso il prodotto scalare:

$$e_{B,A}^{D2}(i) = \langle E(i, j), N_j \rangle^2 \quad (3.7)$$

L'errore Point-to-plane (D2) per l'intera nuvola di punti è quindi definito come la media aritmetica, $e_{B,A}^{D2}$, degli errori $e_{B,A}^{D2}(i)$. Come per PSNRD1, questo errore è espresso come PSNRD2 simmetrico:

$$PSNRD2 = 10 \log_{10} \left(\frac{3 * peak^2}{\max(e_{B,A}^{D2}, e_{A,B}^{D2})} \right) \quad (3.8)$$

3.1.3 Similarità angolare Plane-to-plane

La metrica Plane-to-plane si basa sulla somiglianza angolare dei piani tangenti che corrispondono a punti associati tra un point Cloud in valutazione, B, e una nuvola di punti di riferimento, A. Come per i casi precedenti, vengono individuati i punti corrispondenti tra i due point cloud e per entrambi vengono calcolati i piani tangenti ad ogni punto e le loro normali. Intersecando i piani tangenti corrispondenti, si possono calcolare i due

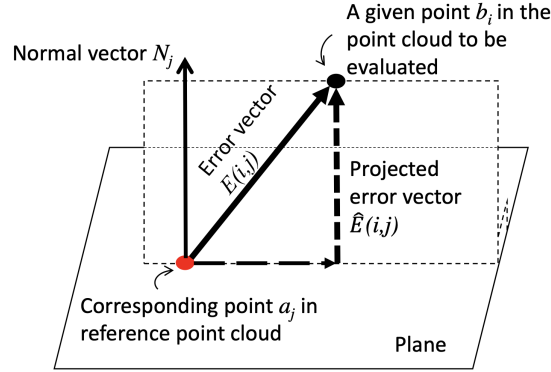


Figura 3.1: Point to plane distance [6]

angoli formati dai vettori normali, il minore dei quali, indicato con θ , viene usato per il calcolo della similarità angolare:

$$Angular\ similarity_{B,A} = 1 - \frac{\theta}{\pi} \quad (3.9)$$

Questa misura, associata ad ogni punto b_j , fornisce un'approssimazione della differenza tra le superfici locali. Usando la MSE tra tutti gli errori locali si può risalire ad una misura totale di distorsione dell'intero point cloud. La somiglianza angolare Plane-to-plane è definita come valore simmetrico, ottenuto dopo aver impostato entrambe le nuvole di punti A e B come riferimento e mantenendo l'errore massimo cioè la somiglianza minima: $\min(Angular\ similarity_{B,A}, Angular\ similarity_{A,B})$.

3.1.4 Metrica Point-to-point per gli attributi

La metrica Point-to-point per gli attributi è basata sull'errore tra i valori degli attributi associati ai punti di un point cloud oggetto della valutazione, B, e del point cloud di riferimento A. I punti associati sono individuati come nei casi precedenti, e il singolo errore è calcolato come differenza tra i valori dell'attributo oggetto di valutazione. Per il colore viene calcolato l'MSE per ogni componente RGB, o YCbCr (che ha una correlazione migliore con la percezione umana) ottenendo poi il PSNR per la singola componente.

$$PSNR = 10 \log_{10} \left(\frac{peak^2}{MSE} \right) \quad (3.10)$$

Il calcolo della PSNR viene effettuato simmetricamente e si seleziona la distorsione massima tra le due. Il valore $peak$ è associato al numero di bit usati per la rappresen-

tazione del colore, nel caso di una profondità colore a 8 bit, il valore *peak* vale 255. Il PSNR colore può essere espresso dalle singole PSNR per i valori Y, Cb e Cr e combinati assieme in un'unica misura di colore:

$$PSNR_{color} = \frac{(6PSNR_y + PSNR_{Cb} + PSNR_{Cr})}{8} \quad (3.11)$$

3.2 Tecniche di valutazione soggettiva

Le tecniche di valutazione soggettive raccolgono le valutazioni di osservatori umani sulla qualità del contenuto o dell'esperienza in esame. L'aggregazione e l'analisi delle singole opinioni costituisce il modo più affidabile per misurare la qualità dell'esperienza dei servizi multimediali. Come è già stato detto in precedenza, le metriche oggettive risulteranno tanto più attendibili ed accurate quanto più riusciranno a coincidere con le valutazioni soggettive. Il giudizio personale è importante per individuare i fattori che più influiscono sulla degradazione della qualità del contenuto e permette di tarare, raffinare e migliorare le tecnologie o gli algoritmi che in qualche modo apportano modifiche al contenuto originale, come strumenti di compressione, trasmissione e supporti di visualizzazione. La qualità soggettiva viene comunemente espressa come un punteggio Mean Opinion Score (MOS) che rappresenta il grado di qualità attribuito da un osservatore medio standard a un dato stimolo. I punteggi MOS vengono raccolti seguendo metodi ben definiti che mirano a garantire impostazioni e condizioni sperimentali identiche, garantendo così un esperimento riproducibile e assicurandone l'accuratezza. I metodi più usati sono raccolti e descritti dalla International Telecommunication Union nello standard ITU-R BT.500-13 [7] per immagini e video e nello standard ITU-R BT.2021 [8] per televisori 3D stereoscopici. Per la valutazione di oggetti e scene 3D, ed in particolare per i pont cloud, non esiste uno standard ad hoc, tuttavia le indicazioni citate in precedenza vengono comunemente adottate anche per la valutazione di oggetti 3D, con opportuni adattamenti. Le metodologie di valutazione soggettiva possono essere a singolo stimolo (SS) o ad doppio stimolo (DS). Nei metodi SS il giudizio ha un valore assoluto, lo stimolo viene giudicato senza avere un riferimento originale e il voto che viene espresso indica il livello di gradimento dell'esperienza che l'utente percepisce realmente in quell'istante secondo il suo giudizio assoluto. Le valutazioni non vengono associate ad un'esperienza di riferimento per quel determinato stimolo, ma dalla sola esperienza pregressa generale dell'utente. Nei metodi DS, invece, il giudizio ha un valore relativo, lo stimolo viene giudicato in relazione ad un contenuto di riferimento che rappresenta il miglior scenario possibile. Il voto espresso

indica quindi un livello di gradimento che l'utente percepisce in relazione all'esperienza pregressa che le è stata indicata come la migliore possibile. Per entrambe le categorie esistono diverse procedure utilizzabili, le più utilizzate sono riassunte nei paragrafi successivi.

3.2.1 Single Stimulus SS

Nei metodi SS, viene presentata una singola immagine, sequenza o nel caso specifico un Point Cloud alla volta e il valutatore fornisce un voto alla presentazione. Il materiale di prova potrebbe includere solo i contenuti degradati sotto studio (Absolute category rating ACR) o potrebbe includere sia i contenuti di prova che quelli di riferimento. In quest'ultimo caso, la sequenza di riferimento viene presentata come uno stimolo indipendente come qualsiasi altro stimolo e il soggetto giudicante non è a conoscenza di quale tra gli stimoli presentati sia l'originale (Absolute category rating with hidden reference ACR-H). Tipicamente i voti sono espressi attraverso una scala ben definita di valori numerici o aggettivi che indicano i livelli di qualità o degradazione dell'immagine: È possibile

ITU-R quality and impairment scales

Five-grade scale			
Quality		Impairment	
5	Excellent	5	Imperceptible
4	Good	4	Perceptible, but not annoying
3	Fair	3	Slightly annoying
2	Poor	2	Annoying
1	Bad	1	Very annoying

Tabella 3.1: ITU-R Five grade scale

usare anche altri metodi di valutazione, con più categorie o livelli numerici (ad esempio un range di valori continuo). Per i contenuti in cui la qualità può fluttuare notevolmente in base alla scena mostrata si può usare il metodo Single stimulus continuous quality evaluation (SSCQE) in cui uno slider elettronico manuale è collegato a un computer ed è utilizzato per registrare la valutazione in modo continuo nel tempo.

3.2.2 Double Stimulus DS

Nei metodi DS, vengono presentate entrambi gli stimoli di test e di riferimento ed il valutatore fornisce un voto al grado di degradazione del materiale sotto osservazione in

relazione a quello di riferimento. Nella metodologia DSIS (double-stimulus impairment scale) al soggetto votante viene prima presentato il modello originale e poi la stessa immagine (o Point Cloud) con un certo livello di degradazione. Successivamente, gli viene chiesto di esprimere un voto sul secondo stimolo ricordando il primo. Per ogni livello di degradazione si forma una coppia riferimento/test, ognuna delle quali viene mostrata al soggetto in ordine casuale. Il metodo DSCQS (double-stimulus continuous quality-scale) chiede al soggetto di valutare due versioni dello stesso stimolo mostrate contemporaneamente uno a fianco all'altro, una dei quali è soggetta a degradazione e l'altra è inalterata. La versione originale serve da riferimento ma al valutatore non viene comunicato quale delle due sia, e la sua posizione viene cambiata in ordine casuale per ogni coppia di stimoli. Per valutare la fedeltà dei contenuti che variano nel tempo viene poi usato il metodo Simultaneous double stimulus for continuous evaluation (SDSCE), in cui le due sequenze sono mostrate fianco a fianco contemporaneamente e al soggetto è chiesto di valutare la somiglianza tra le due tramite uno slider elettronico manuale. La sequenza di riferimento è nota al valutatore.

Per entrambe le categorie di valutazione soggettiva, gli stimoli possono rappresentare i point cloud, visualizzati tramite uno schermo, un visore per la realtà virtuale o altri sistemi.

3.3 Analisi dei dati

Nel corso di un esperimento soggettivo tipicamente vengono raccolte grandi quantità di dati. Questi dati devono essere condensati da tecniche statistiche per produrre risultati in forma grafica e numerica che riassumano le prestazioni dei sistemi in prova. La misura statistica più importante è la Mean Opinion Score MOS che, come detto in precedenza, rappresenta il grado di qualità attribuito da un osservatore medio standard a un dato stimolo.

$$MOS = \frac{1}{N} \sum_{i=1}^N score_{i,j,k,r} \quad \forall j, k, r \quad (3.12)$$

Dove N sono i soggetti partecipanti, j sono gli stimoli, k le varie condizioni ed r le ripetizioni. I punteggi medi vengono accompagnati da un intervallo di confidenza che deriva dalla deviazione standard e dalle dimensioni di ciascun campione. La deviazione standard è così definita:

$$\sigma = \sqrt{\frac{1}{M} \sum_{i=1}^M (score_i - m)^2} \quad (3.13)$$

con M =numero di sample, m =media (MOS); mentre lo standard error:

$$SE = \frac{\sigma}{\sqrt{M}} \quad (3.14)$$

In letteratura viene proposto come standard un intervallo di confidenza del 95% che è calcolato attraverso il valore ci derivato da SE :

$$ci = 1,96 \cdot SE \quad (3.15)$$

che si traduce per i punteggi MOS:

$$[MOS - ci; MOS + ci] \quad (3.16)$$

Capitolo 4

Esperimento soggettivo

4.1 Obiettivi

Come illustrato nei capitoli precedenti i test soggettivi della qualità sono importanti per raccogliere una valutazione “reale” dell’esperienza in esame, che rispecchi la percezione di un utente medio. I giudizi raccolti forniscono un punto di partenza per lo sviluppo e il miglioramento delle tecnologie coinvolte e permettono di trovare il trade-off migliore tra gli obiettivi di funzionalità prefissati e i requisiti qualitativi richiesti. L’obiettivo di questo lavoro di tesi è raccogliere e analizzare dati soggettivi relativi alla QoE nella visualizzazione di Point Cloud compressi in ambiente VR, con particolare attenzione alla geometria e alla struttura dei modelli. Il compito dei partecipanti è infatti giudicare la componente visiva geometrica dei PC, facendo attenzione alla presenza di artefatti quali discontinuità nelle facce della mesh, sovrapposizioni o mal posizionamenti degli elementi costituenti l’oggetto, deformazioni nella forma e degradazioni dei dettagli.

4.2 Partecipanti

In questo studio viene coinvolto un campione composto da 17 partecipanti, di età compresa tra 23 e 62 anni, di cui 11 maschi e 6 femmine. Per una prima fase di collaudo e messa a punto dell’esperimento, sono coinvolti 2 ulteriori soggetti, ai quali è stato affidato il compito di fornire un feedback sull’intera esperienza. I dati associati a questi due partecipanti non sono inclusi nell’analisi dei risultati.

4.3 Apparatus

Come supporto di visualizzazione si è scelto l'Oculus Quest 2, un visore VR HMD standalone a 6 gradi di libertà. Queste caratteristiche sono particolarmente rilevanti in quanto garantiscono elevata libertà di movimento e la possibilità di osservare i modelli Point Cloud da differenti distanze, altezze e angolazioni. La portabilità e l'indipendenza dell'Oculus facilitano inoltre la preparazione dell'esperimento, così come la sua esecuzione. Questo visore è prodotto da Meta (Facebook) e monta un sistema operativo Android, integra il tracciamento delle mani e dell'ambiente circostante grazie a 4 telecamere poste nella parte frontale; lo schermo LCD ha una risoluzione di 1832 x 1920 pixel per occhio con campo visivo di 90° e supporta refresh rate di 60, 72 e 90 Hz (72 Hz di default).



Figura 4.1: Oculus Quest 2 by Meta

4.4 Dataset

4.4.1 Selezione dei modelli

Per comporre il dataset sono stati selezionati 4 modelli Point Cloud, rappresentanti oggetti o scenari afferenti ad ambiti diversi:

- *Abbey*: veduta aerea delle rovine della cattedrale di Fountains Abbey e dintorni.
- *Liddedewer*: recipiente con coperchio appartenente al patrimonio artistico del museo nazionale di arti asiatiche dello Smithsonian.
- *Mercedes*: automobile Mercedes Benz GLS 580

- *Valley*: veduta aerea di un paesaggio montuoso.

Un ulteriore modello *Specchiera* è stato scelto come oggetto esemplificativo per la fase introduttiva di training, durante la quale ai soggetti viene concesso di prendere confidenza con l'ambiente virtuale e di comprendere il compito a loro assegnato. Per poter giudicare al meglio la sola geometria dei modelli, texture e componente cromatica sono state rimosse mantenendo tutti i point cloud di colore bianco uniforme. Inoltre, per non perdere la percezione di tridimensionalità e verosimiglianza sono presenti durante la visualizzazione le ombre che l'oggetto proietta su sé stesso causate da protuberanze, rilievi e forma. Il modello *Palacio*, che in origine doveva comparire nel dataset, è stato escluso perché di dimensioni troppo elevate, che causavano il collasso dei vertici durante lo scaling a dimensioni inferiori.

4.4.2 Decimazioni

Per ogni modello è stato effettuato un processo di degradazione geometrica utilizzando il software open source di grafica tridimensionale Blender. Questa operazione è stata realizzata applicando il modificatore *decimate* che permette di ridurre il numero di facce e vertici di una mesh causando il minor cambiamento possibile nella forma dell'oggetto. Attraverso le opzioni *collapse* e *triangulate* il processo di decimazione funziona attraverso la fusione progressiva dei vertici e la conseguente generazione delle facce attraverso la triangolazione degli stessi. Il valore del parametro Ratio indica il rapporto di facce/vertici da mantenere dopo tale elaborazione: al valore 1.0 la mesh rimane inalterata, al valore 0.5 i vertici vengono fusi in modo tale che rimanga la metà del numero di facce, al valore 0 tutte le facce sono rimosse. Per ogni point cloud sono state create 5 versioni degradate impostando valori di Ratio pari a 0.5, 0.2, 0.1, 0.005, 0.001 corrispondenti a livelli di decimazioni del 50%, 20%, 10%, 5% e 1%. Per alcuni modelli, i livelli più bassi di ratio hanno causato la degenerazione dell'oggetto, con perdita di facce e contorni. In particolare, i Point cloud *Mercedes* ed *Abbey* per il livello 0.001 di ratio sono stati esclusi dal dataset perché troppo degradati.

4.4.3 Dataset finale

Nel dataset finale si è scelto di inserire anche i modelli originali inalterati. Per ogni point cloud risultano quindi utilizzate 6 versioni differenti, ad eccezione dei modelli *Abbey* e *Mercedes* per i quali le versioni decimate a 1% sono state scartate. Gli stimoli utilizzati

nell'esperimento sono in totale 22, più i 3 stimoli di train, per un numero complessivo di 25 modelli visualizzati per ogni sessione.

4.5 Esperimento

4.5.1 Protocollo

Il protocollo seguito per l'esperimento è l'ACR-H, con riferimento alle raccomandazioni ITU-R BT.500-13. Ai partecipanti non viene fornita nessuna informazione riguardo a quale tra i PC sia l'originale né al livello di degradazione associato ai modelli. Gli stimoli, inoltre, vengono presentati ad ogni sessione in ordine casuale, al fine di ridurre il più possibile qualsiasi bias dovuto alla visualizzazione ordinata dei modelli e dalla conoscenza pregressa dei vari livelli di degradazione. L'obiettivo di questo esperimento è infatti quello di raccogliere dei giudizi assoluti, scollegati da una percezione della qualità basata sul confronto con uno stimolo di riferimento.

4.5.2 Ambiente virtuale

Il software utilizzato per condurre l'esperimento è stato realizzato in Unity e programmato in linguaggio C#. Sono state apportate alcune modifiche nelle impostazioni del progetto originale per adattare l'applicazione alle esigenze dell'esperimento. L'ambiente virtuale è costituito da una stanza quadrata spoglia e di colore bianco uniforme, di dimensioni 10x10x7 metri, al centro della quale è presente un cubo di lato 1 metro con funzione di piedistallo per il posizionamento degli stimoli. I PC sono stati scalati utilizzando questo piedistallo come riferimento per le dimensioni e posizionati in modo che apparissero fluttuanti appena sopra di esso. La superficie di camminamento reale necessaria per le osservazioni si è dimostrata nettamente inferiore a quella della stanza virtuale, ma abbastanza estesa da consentire il movimento attorno al piedistallo centrale. Le misure ideali si sono rivelate non inferiori a 4x4 metri. Per ottimizzare i fattori descritti in precedenza, quali dimensione e posizione degli stimoli, così come spazio di movimento e durata di visualizzazione di ogni singolo modello, sono state condotte due sessioni preliminari, già citate in precedenza, durante le quali è stato chiesto ai partecipanti di fornire un feedback su questi aspetti.

4.5.3 Esecuzione

L'esperimento è suddiviso in due fasi principali, una di training e una di testing. Durante la fase di training, il soggetto ha modo di prendere confidenza con l'ambiente, con la visualizzazione dei modelli e la possibilità di movimento, così come di sistemare il visore se non indossato correttamente, e di comprendere a pieno il compito assegnatogli. Il modello utilizzato per questa parte della sessione è il modello "specchiera" fornito in 3 versioni corrispondenti ai livelli di decimazioni 100%, 50% e 1%. Ai partecipanti è richiesto di assegnare un voto anche ai modelli di training, in modo da fornire una breve esperienza esemplificativa del funzionamento del test e dei livelli di degradazione utilizzati. I dati raccolti in questa fase non sono considerati parte dell'esperimento.



Figura 4.2: Fase di training, modello *Specchiera*

Il test comincia immediatamente dopo la visualizzazione dei tre modelli iniziali. Ai partecipanti viene chiesto di concentrarsi sulla qualità percepita per ogni stimolo, e fornire un voto secondo la propria sensibilità e i propri canoni. Ai partecipanti non viene posta limitazione alla possibilità di movimento. Infatti, si consiglia loro di muoversi attorno allo stimolo per apprezzare al meglio i dettagli, potendo osservare i PC da più angolazioni e da distanze diverse. Per chi lo ritenesse opportuno viene concesso anche di passare

4. Esperimento soggettivo

attraverso i modelli. Il processo di valutazione per ogni PC si suddivide in due parti: osservazione e valutazione. Ogni modello viene presentato singolarmente per 18 secondi durante i quali il soggetto può osservare il PC.



Figura 4.3: Fase di test, modello *Lidded ewer*

Terminato il tempo di osservazione, è richiesto di fornire un voto al modello appena osservato, esprimendo un giudizio sulla qualità in una scala da 1 a 5: 1-Pessima, 2-Bassa, 3-Discreta, 4-Buona, 5- Eccellente. La raccolta delle valutazioni avviene tramite un'apposita interfaccia posta sul piedistallo al posto dello stimolo, che consente di selezionare il voto scelto tramite i controller dell'Oculus Quest 2. Non è presente nessuna limitazione di tempo per questa fase.

L'intero esperimento è quindi strutturato in modo tale che dopo ogni periodo di osservazione il soggetto possa immediatamente fornire il giudizio sullo stimolo appena osservato. L'alternanza osservazione/valutazione viene ripetuta per ogni modello e la durata media dell'intero esperimento è di circa 10 minuti.

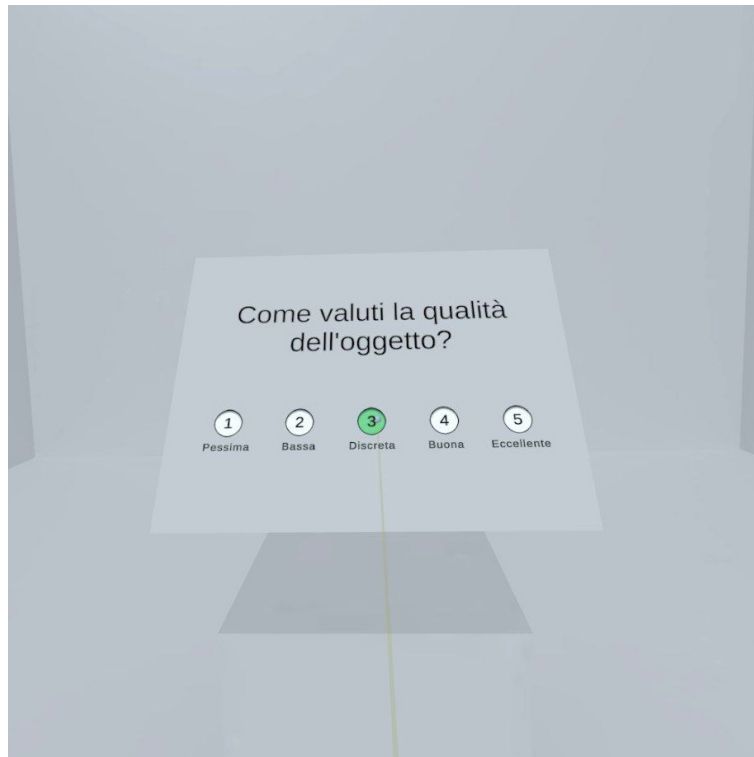


Figura 4.4: Fase di votazione

4.5.4 Osservazioni

Il tempo di visualizzazione, inizialmente impostato a 15 secondi, è stato esteso a 18 secondi dopo avere condotto le due sessioni preliminari di tuning, in cui i soggetti hanno espresso l'esigenza di osservare i modelli con più calma. Allo stesso modo, la risposta dei soggetti ha permesso di regolare la posizione dei PC sopra il piedistallo. I modelli rappresentanti paesaggi e vedute aeree sono stati posizionati ad una altezza inferiore rispetto ai PC rappresentanti oggetti, conseguenza del fatto che la dimensione in altezza dei modelli paesaggistici è nettamente inferiore rispetto alle altre e la maggior parte dei dettagli vengono apprezzati attraverso una osservazione dall'alto.

4. Esperimento soggettivo



Figura 4.5: Abbey 1% (scartato)

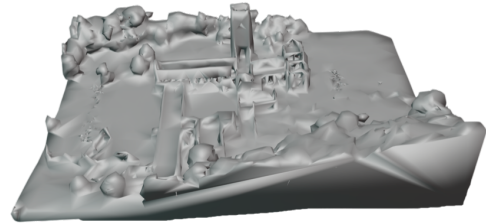


Figura 4.6: Abbey 5%

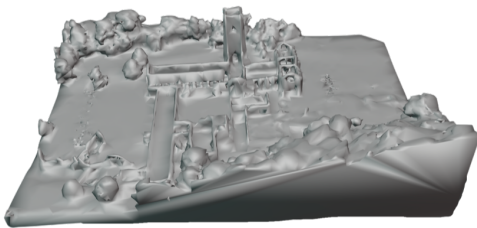


Figura 4.7: Abbey 10%

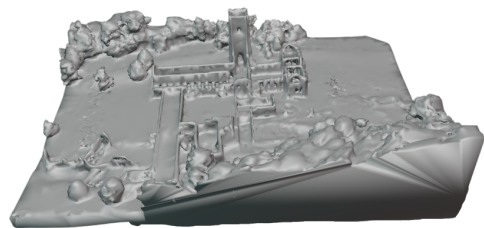


Figura 4.8: Abbey 20%

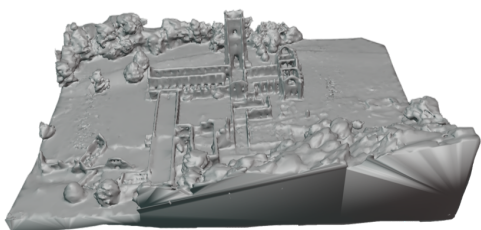


Figura 4.9: Abbey 50%

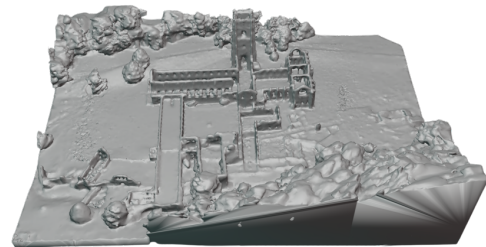


Figura 4.10: Abbey 100%

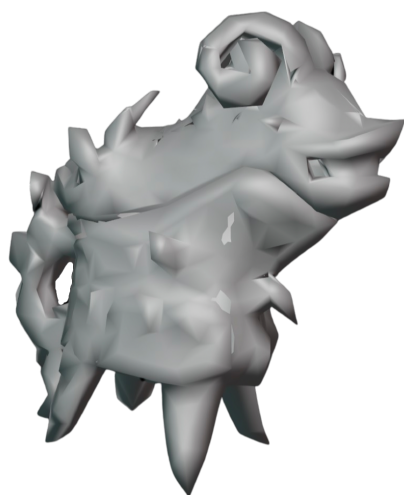


Figura 4.11: Lidded ewer 1%



Figura 4.12: Lidded ewer 5%



Figura 4.13: Lidded ewer 10%



Figura 4.14: Lidded ewer 20%



Figura 4.15: Lidded ewer 50%



Figura 4.16: Lidded ewer 100%

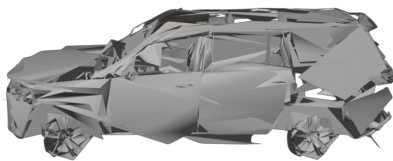


Figura 4.17: Mercedes 1% (scartato)



Figura 4.18: Mercedes 5%



Figura 4.19: Mercedes 10%



Figura 4.20: Mercedes 20%



Figura 4.21: Mercedes 50%



Figura 4.22: Mercedes 100%

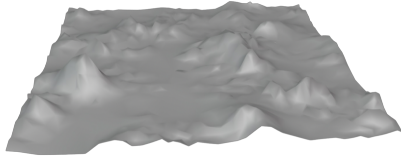


Figura 4.23: Valley 1%

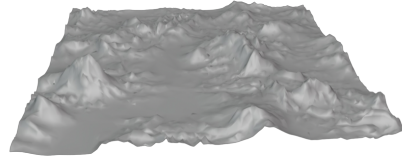


Figura 4.24: Valley 5%

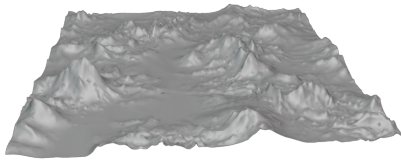


Figura 4.25: Valley 10%

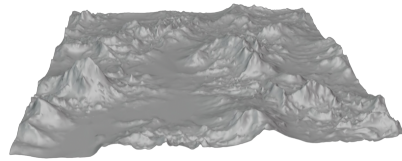


Figura 4.26: Valley 20%

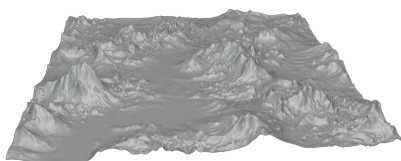


Figura 4.27: Valley 50%

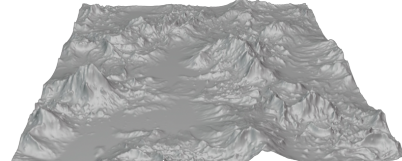


Figura 4.28: Valley 100%

Capitolo 5

Analisi dei dati

5.1 Dati raccolti

I dati relativi all'esperimento soggettivo sono raccolti durante l'esecuzione del test, e registrati in appositi file direttamente dall'applicazione grazie ad opportuni script. Le informazioni memorizzate sono molteplici, ma solo alcune di queste vengono utilizzate in questo studio. Dopo ogni fase di osservazione, il software genera un file di Log contenente, per ogni frame, i dati relativi alla posizione occupata dall'utente, l'orientamento del visore e il framerate puntuale. In questo file sono anche indicati l'ID dell'utente, l'ID dello stimolo e il timestamp. Al termine dell'intera sessione, l'applicazione genera un ulteriore file in formato .csv relativo alle valutazioni espresse per ogni PC, strutturato per colonne in maniera seguente:

- **ID utente:** codice alfanumerico pseudocasuale che rappresenta in forma anonima il soggetto che ha effettuato il test
- **ID del modello:** stringa composta dalla concatenazione del nome del modello (e.g. *abbey*) e del grado di decimazione (e.g. 005) separati dal carattere underscore (e.g. *abbey_005*). Per i modelli originali il grado di decimazione viene indicato con HQ (e.g. *abbey_HQ*).
- **voto espresso:** valore numerico da 1 a 5.
- **timestamp** di inizio e fine della scena espresso in millisecondi.

Come è stato illustrato nei capitoli precedenti, l'obiettivo di questa tesi è raccogliere dati soggettivi sulla qualità visiva dei point cloud dal punto di vista geometrico e strutturale.

Per questo motivo le informazioni utilizzate in questo progetto sono quelle salvate nel file .csv contenente i giudizi espressi dai partecipanti. Le informazioni aggiuntive raccolte permettono, tuttavia, di esaminare altri aspetti significativi dell'esperimento. Si potrebbe ad esempio condurre un'analisi dei pattern comportamentali in ambiente VR o sull'effetto del calo di framerate sulla percezione della qualità. Approfondimenti di questo tipo, ed ulteriori elaborazioni in ambito IA sono stati trattati da Luca Hartz [9] e da Francesco Pham [10] nelle corrispettive tesi magistrali.

5.2 Operazioni preliminari: outliers

Come operazione preliminare all'analisi statistica dei dati, è stato effettuato uno screening dei partecipanti sulla base delle valutazioni espresse. La procedura seguita è quella raccomandata nel documento ITU-R BT.500-13 Annex 2, ed è stata effettuata utilizzando il software Matlab. La funzione di screening non ha rilevato outliers, nessun soggetto è stato quindi scartato.

5.3 Mean Opinion Score: MOS

Per effettuare il calcolo dei punteggi MOS, le valutazioni degli utenti sono raggruppate e suddivise secondo dei criteri ben precisi:

- per livello di dettaglio per singolo PC
- per livello di dettaglio complessivo

5.3.1 MOS per Point Cloud

Per ogni Point Cloud, e per tutte le rispettive versioni, sono calcolati i punteggi MOS, deviazione standard e intervallo di confidenza.

Abbey

In figura 5.1 sono riportati i valori di MOS relativi al Point Cloud *Abbey*. È interessante notare come il modello originale non sia stato affatto riconosciuto come tale. Il voto medio per la versione corrispondente ad una decimazione al 50% è infatti maggiore. Si può affermare che il PC decimato al 50% è intercambiabile con quello originale. Un risultato di

questo tipo può essere rilevante se si pensa ad applicazioni in cui è fondamentale trasmettere i modelli velocemente, in cui le esigenze di occupazione di memoria sono stringenti, o in cui si ha a disposizione una ridotta capacità di calcolo, ma è comunque richiesta un'elevata qualità di visualizzazione. Infatti, se l'utente finale non si accorge della differenza tra originale e versione degradata, risulta più efficiente trasmettere, memorizzare o renderizzare la versione decimata, in quanto meno dispendiosa dal punto di vista delle risorse. Gli intervalli di confidenza, in particolare per le versioni molto degradate, sono più ampi se confrontati con i corrispettivi intervalli negli altri modelli.

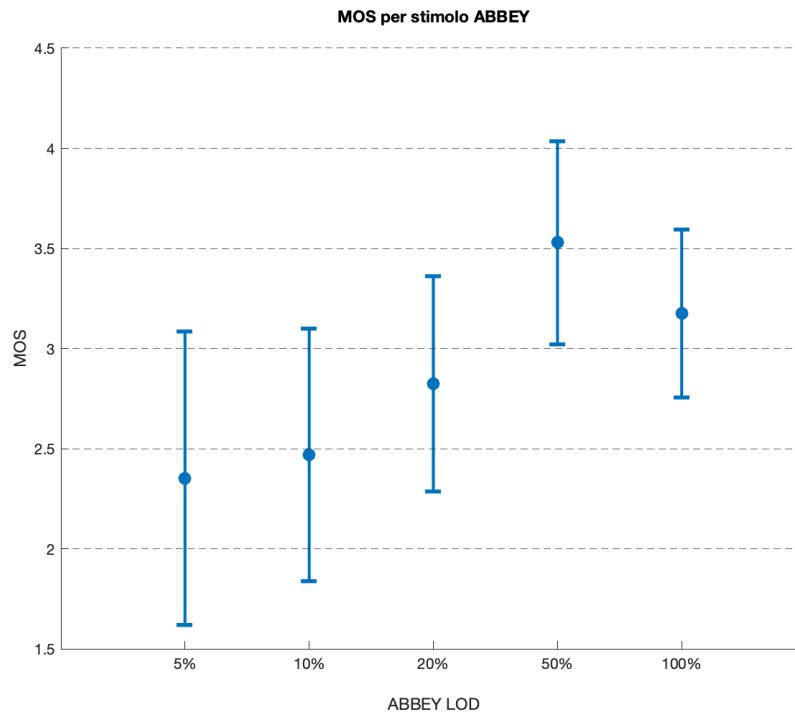


Figura 5.1: Valori dei MOS per il modello Abbey

	ABBEY 5%	ABBEY 10%	ABBEY 20%	ABBEY 50%	ABBEY 100%
MEDIA	2.35	2.47	2.82	3.53	3.18
DEV.STANDARD	1.54	1.33	1.13	1.07	0.88

Tabella 5.1: Valori di MOS e deviazione standard per il modello Abbey

Lidded ewer

In figura 5.2 sono riportati i valori di MOS relativi al Point Cloud *Lidded ewer*. Le considerazioni fatte per il modello precedente non valgono per questo Point Cloud. I voti medi per ogni grado di decimazione seguono un andamento abbastanza lineare: la qualità

5. Analisi dei dati

percepita aumenta al crescere del LOD (Level Of Detail) con differenze di voto regolari. Solo al di sotto del livello 5% si ha un notevole peggioramento nella valutazione della qualità. Gli intervalli di confidenza sono inoltre ristretti rispetto agli altri modelli.

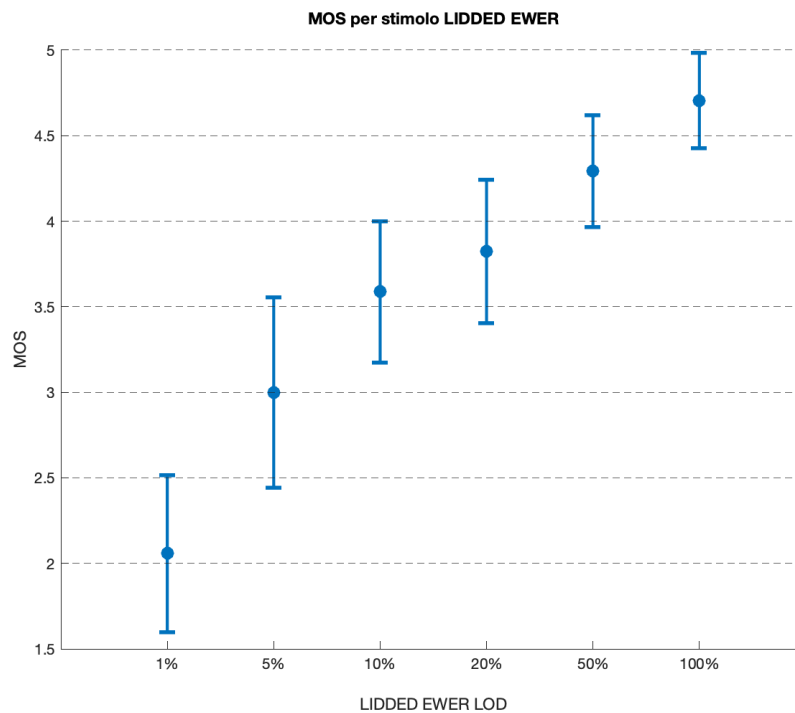


Figura 5.2: Valori dei MOS per il modello Lidded ewer

	LIDDEDEWER 1%	LIDDEDEWER 5%	LIDDEDEWER 10%	LIDDEDEWER 20%	LIDDEDEWER 50%	LIDDEDEWER 100%
MEDIA	2.06	3.00	3.59	3.82	4.30	4.71
DEV.STANDARD	0.97	1.17	0.87	0.88	0.69	0.59

Tabella 5.2: Valori di MOS e deviazione standard per il modello Lidded ewer

Mercedes

In figura 5.3 sono riportati i valori di MOS relativi al Point Cloud *Mercedes*. L'andamento dei punteggi è molto simile a quello del modello *Lidded ewer*, le differenze di voto tra LOD si mantengono pressoché costanti, mentre gli intervalli di confidenza risultano in media più estesi rispetto al Point Cloud precedente, specialmente se ci si avvicina ai livelli in cui è stata apportata una degradazione più spinta. Si può notare, inoltre, come i valori di MOS siano uniformemente distribuiti su quasi tutta la scala dei voti.

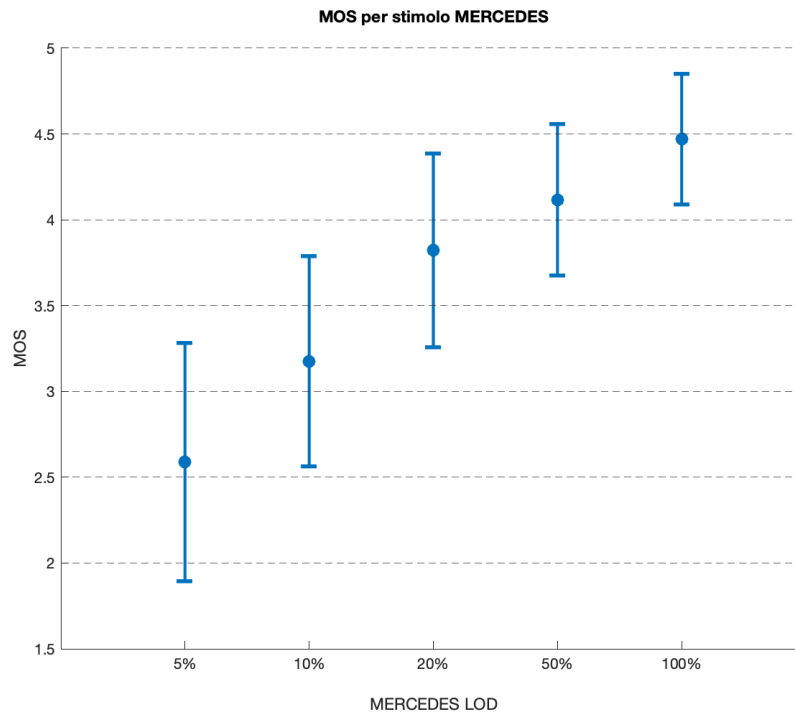


Figura 5.3: Valori dei MOS per il modello Mercedes

	MERCEDES 5%	MERCEDES 10%	MERCEDES 20%	MERCEDES 50%	MERCEDES 100%
MEDIA	2.59	3.18	3.82	4.12	4.47
DEV.STANDARD	1.46	1.29	1.19	0.93	0.80

Tabella 5.3: Valori di MOS e deviazione standard per il modello Mercedes

Valley

In figura 5.4 sono riportati i valori di MOS relativi al Point Cloud *Valley*. L'andamento dei punteggi MOS raccolti per questo modello si discostano da tutti gli altri. Si può notare come i MOS per il livello 10% e 20% siano relativamente elevati se rapportati ai modelli più definiti. Questi due livelli sono inoltre perfettamente sovrapponibili, medie e varianze risultano infatti identiche. Dal livello 5% in giù, si può apprezzare invece un netto calo nei punteggi. Il dato più interessante si ha tuttavia in corrispondenza del livello 100%. Il Point Cloud originale, infatti, viene considerato in media peggiore o alla pari delle sue versioni degradate, condizione che si verifica per i livelli 50%, 20% e 10%. A differenza degli altri LOD l'intervallo di confidenza risulta anche più esteso. Questo fenomeno è dovuto al fatto che lo stimolo *valley_HQ* è l'unico tra quelli presentati che causa un drastico decremento del framerate durante la sua visualizzazione. Al fine di verificare questa ipotesi, è stata calcolata la media dei frame per secondo durante l'osservazione di ogni stimolo. Come si può notare in figura 5.9, la media degli fps per lo stimolo *valley_HQ*

5. Analisi dei dati

è pari a 59.23 mentre per tutti gli altri modelli è ben superiore ai 70. Si ricorda che ai partecipanti è stato chiesto di valutare la qualità dei PC osservando la geometria e il livello di dettaglio strutturale degli stimoli. Una forte riduzione nel framerate è pertanto un fattore rilevante nella percezione della qualità visiva generale.

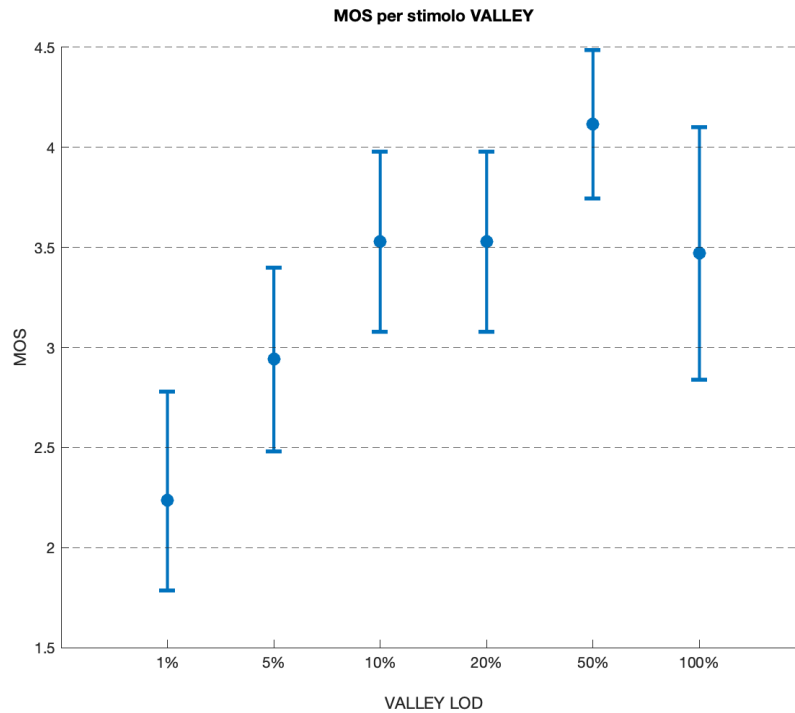


Figura 5.4: Valori dei MOS per il modello Valley

	VALLEY 1%	VALLEY 5%	VALLEY 10%	VALLEY 20%	VALLEY 50%	VALLEY 100%
MEDIA	2.24	2.94	3.53	3.53	4.12	3.47
DEV.STANDARD	1.15	0.97	0.94	0.94	0.78	1.33

Tabella 5.4: Valori di MOS e deviazione standard per il modello Valley

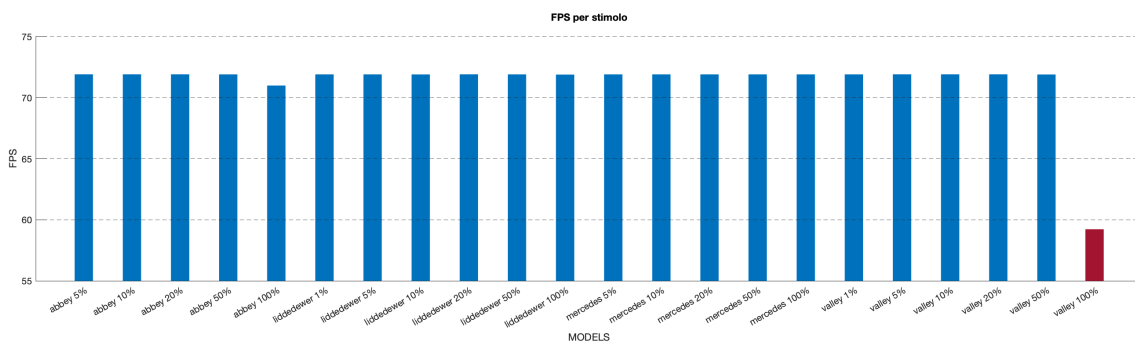


Figura 5.5: Valore medio del framerate per ogni stimolo

MOS per tutti gli stimoli

Confrontando i punteggi MOS di Point Cloud diversi si possono notare alcuni aspetti legati alla natura del singolo PC. I modelli paesaggistici hanno punteggi tendenzialmente meno elevati rispetto ai modelli rappresentanti oggetti. Questo fenomeno può essere dovuto al fatto che la rappresentazione paesaggistica ad ampio raggio, che per sua natura comprende un elevato numero di informazioni, non sia in grado di fornire lo stesso grado di dettaglio che si può trovare in un modello di un oggetto, in cui le informazioni da rappresentare sono inferiori. Il modello *Lidded ewer* raffigura un manufatto museale ricco di dettagli, ben distinguibili nella versione più definita. Il modello *Abbey* invece, pur rappresentando le intere rovine della cattedrale di Fountains Abbey e i suoi dintorni, elementi che nella realtà contengono molti più dettagli del singolo oggetto museale, non è in grado di trasmettere con la stessa precisione le informazioni. Gli alberi, ad esempio, sono riconoscibili solo in relazione al contesto, se separati dalla scena non sono distinguibili. I modelli paesaggistici forniscono quindi un'approssimazione di ciò che rappresentano, condizione che influenza la percezione della qualità generale del modello in esame.

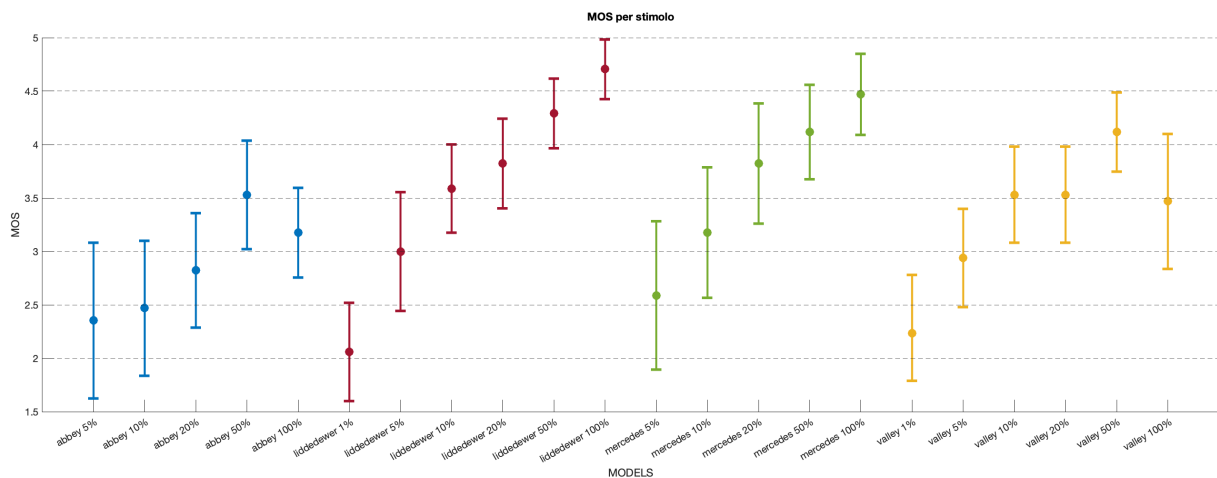


Figura 5.6: Valori di MOS per tutti gli stimoli

5.3.2 MOS per livello di dettaglio

Per ogni livello di dettaglio LOD sono calcolati i punteggi MOS, deviazione standard e intervallo di confidenza. In questa analisi vengono raggruppate le versioni degradate dei modelli a cui è stato applicato lo stesso livello di decimazione. Si può notare come i punteggi MOS siano coerenti con i gradi di decimazione, all'aumentare del livello di dettaglio aumenta la valutazione media della qualità percepita. Tuttavia, come già osservato in

precedenza, gli stimoli appartenenti ai livelli 50% e 100% risultano in media valutati di eguale qualità, e quindi non facilmente distinguibili. L'osservazione fatta per il modello *Abbey* è quindi generalizzabile. Se l'utente finale non si accorge della differenza tra modello originale e versione degradata, risulta più efficiente utilizzare la seconda in quanto meno dispendiosa in termini di occupazione di memoria, capacità trasmissiva e potenza computazionale necessaria per il rendering. Applicazioni real-time possono ad esempio beneficiare dell'intercambiabilità dei modelli, ed utilizzare versioni decimate più facilmente trasmissibili e renderizzabili. Il modello *Abbey*, nella sua versione originale, necessita di uno spazio di memorizzazione pari a 30 MB, contro i 5 MB per la versione degradata al 50%. In questo caso, a parità di qualità percepita, si registra per l'occupazione in memoria una riduzione superiore all'80%. Risulta evidente, quindi, che per applicazioni che hanno a disposizione un numero ridotto di risorse, la possibilità di utilizzare versioni decimate di Point Cloud sia un enorme vantaggio. Va ricordato che per il LOD 100%, i giudizi raccolti sul modello *Valley*, affetto da calo del framerate, influiscono a ribasso sulla media. È rilevante evidenziare il fatto che è ben osservabile la just-noticeable difference, JND, in corrispondenza del salto tra il livello 50% e il livello 20%.

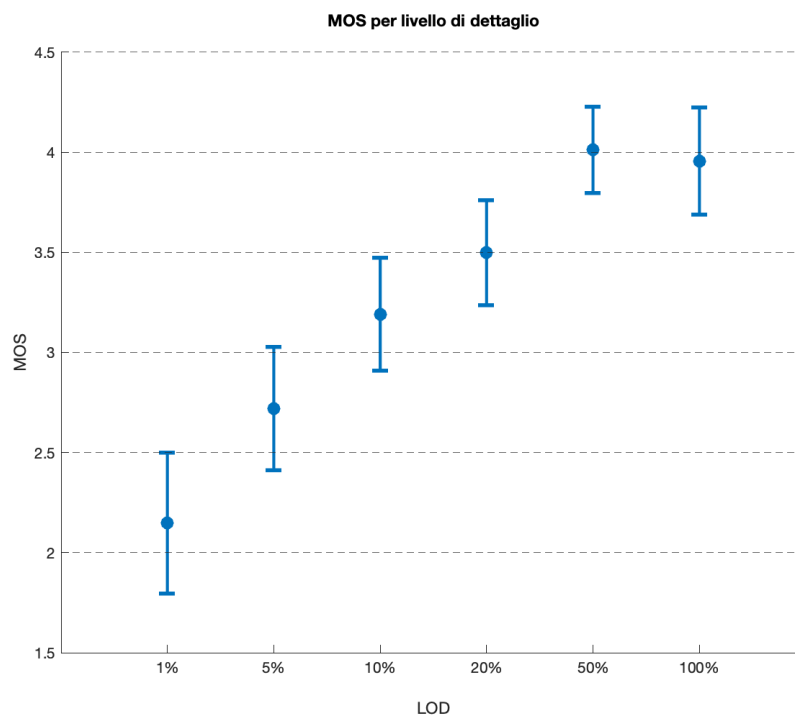


Figura 5.7: Valori di MOS per livello di dettaglio

5.4 T-Test

Un'ulteriore indagine statistica, il T-Test, è stata effettuata sui dati aggregati per livello di dettaglio. Questo test permette di confermare o smentire l'ipotesi nulla che non ci sia differenza statistica significativa tra gruppi di valutazioni appartenenti a LOD diversi. Il test è stato eseguito in Matlab utilizzando la funzione $ttest(x,y)$, che riporta in output il valore 1 se l'ipotesi nulla viene rigettata, 0 altrimenti. Come ci si poteva aspettare, i LOD 50% e 100% non risultano significativamente indipendenti, le medie e gli intervalli di confidenza sono quasi totalmente sovrapponibili. Un risultato simile si può osservare per i livelli 10% e 20%. Tutte le altre coppie di distribuzioni risultano significativamente indipendenti.

LOD	T-TEST
5% - 10%	1
5% - 20%	1
5% - 50%	1
5% - 100%	1
10% - 20%	0
10% - 50%	1
10% - 100%	1
20% - 50%	1
20% - 100%	1
50% - 100%	0

Tabella 5.5: Risultati del T-Test

5.5 Dati secondari

Di seguito sono riportati dati secondari ed osservazioni raccolte durante e dopo l'esperimento soggettivo:

- La QoE generale è risultata inferiore in corrispondenza della visualizzazione del modello *Valley_HQ*. Molti partecipanti hanno riportato questo commento alla fine dell'esperimento.
- La maggior parte dei soggetti ha espresso la volontà, se fosse stato possibile, di cambiare i voti già assegnati dopo aver osservato versioni migliori o più degradate dello stesso modello solo in seguito alla votazione.

5. Analisi dei dati

- Un'osservazione comune riguarda il fatto che tanti modelli fossero in apparenza uguali, causando difficoltà nel dare un voto senza conoscere il modello di riferimento o il livello di degradazione corrispondente.
- Tra tutti i partecipanti, uno solo ha dichiarato di aver accusato fastidio fisico, in particolare una leggera sensazione di nausea.
- Quasi la totalità dei soggetti non si era mai approcciata alla realtà virtuale, solo 4 persone avevano esperienza pregressa.
- Si è riscontrata una tendente facilità di utilizzo del visore e dei controller, così come una maggior comprensione dei task per i soggetti più giovani.
- Quasi tutti i partecipanti hanno espresso la necessità di abituarsi per qualche secondo sia in "ingresso che in uscita" dal visore.
- Tutti i partecipanti hanno giudicato divertente e piacevole l'esperienza della realtà virtuale con il visore Oculus Quest 2 (parere espresso anche dal soggetto che ha accusato nausea).

Conclusioni

L'analisi statistica dei giudizi soggettivi forniti da un campione di utenti rappresenta un ottimo punto di partenza per lo sviluppo e il miglioramento delle tecnologie sotto studio e permette di trovare il trade-off migliore tra gli obiettivi di funzionalità prefissati e i requisiti qualitativi richiesti. In questo lavoro di tesi sono stati registrati ed analizzati dati soggettivi relativi alla QoE nella visualizzazione di Point Cloud compressi in ambiente VR. Per la raccolta dati è stato allestito un esperimento soggettivo, utilizzando come supporto di visualizzazione VR il visore Oculus Quest 2. Si è scelto di seguire il protocollo ACR-H, facendo riferimento alle raccomandazioni ITU-R BT.500-13. Ai partecipanti è stato chiesto di giudicare la componente visiva geometrica dei PC, facendo attenzione alla presenza di artefatti quali discontinuità nelle facce dei modelli, sovrapposizioni o mal posizionamenti degli elementi costituenti gli oggetti, deformazioni nella forma e degradazioni dei dettagli. I dati sono stati poi aggregati e analizzati attraverso diversi indici statistici. In particolare, è stato calcolato il punteggio MOS per ogni stimolo e per ogni livello di dettaglio, assieme a deviazione standard e intervallo di confidenza. Le distribuzioni appartenenti ai diversi LOD sono state inoltre confrontate tramite il T-Test. Lo studio di questi indicatori ha fatto emergere due evidenze rilevanti. Mediamente, i modelli appartenenti ai livelli di dettaglio più elevati non sono stati distinti. I Point Cloud a cui è stato apportato un processo di decimazione fino al 50% possono sostituire in media il modello originale senza conseguenza sulla qualità visiva percepita dall'utente finale. Si nota invece un netto calo di prestazioni visive per gli stimoli decimati al di sotto del 50% pur mantenendo livelli discreti di qualità fino a 10%. Il secondo aspetto significativo riguarda l'effetto del calo di framerate sulla QoE. È stato osservato come questo fenomeno abbia una influenza rilevante non solo sulla qualità dell'esperienza complessiva, ma anche sulla qualità percepita degli stimoli a livello prettamente visivo. L'analisi di queste osservazioni può aiutare a comprendere come il processo di compressione e la conseguente perdita di informazione possa influenzare la qualità percepita da un utente finale in determinate circostanze e come il calo di framerate sia determinante per la resa

CONCLUSIONI

visiva, dando la possibilità di migliorare il già citato trade-off tra funzionalità e qualità.

Bibliografia

- [1] *JPEG Pleno Point Cloud Use Cases and Requirements, v1.6*. URL: https://ds.jpeg.org/documents/jpegpleno/wg1n100096-094-REQ-JPEG_Pleno_Point_Cloud_Use_Cases_and_Requirements_v1.6.pdf.
- [2] Graziosi D. et al. *An overview of ongoing point cloud compression standardization activities: video-based (V-PCC) and geometry-based (G-PCC)*. URL: <https://www.cambridge.org/core/journals/apsipa-transactions-on-signal-and-information-processing/article/an-overview-of-ongoing-point-cloud-compression-standardization-activities-videobased-vpcc-and-geometrybased-gpcc/56FCAF660DD44348BCB1BCA9B5EC56CF>.
- [3] Erica Rocchi, Marco Carli e Federica Battisti. *Visibility and annoyance of artifacts in point cloud images*.
- [4] *JPEG Pleno Point Cloud Coding Common Test Conditions v3.6*. URL: https://ds.jpeg.org/documents/jpegpleno/wg1n91058-PCQ-JPEG_Pleno_Point_Cloud_Common_Test_Conditions_v3_6.pdf.
- [5] Alireza Javaheri et al. *Subjective and Objective Quality Evaluation of Compressed Point Clouds*. URL: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8122239>.
- [6] Dong Tian et al. *geometric distortion metrics for point cloud compression*. URL: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8296925>.
- [7] *Reccommendation ITU-R BT.500-13*. URL: https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.500-13-201201-I!!PDF-E.pdf.
- [8] *ITU-R BT.2021, Subjective methods for the assessment of stereoscopic 3DTV systems*. URL: https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.2021-0-201208-S!!PDF-E.pdf.

Bibliografia

- [9] Luca Hartz. *Ottimizzazione della Qualità Soggettiva in Dispositivi di Realtà Virtuale*.
- [10] Francesco Pham. *Deep 3D model optimization for immersive interactive applications*.

Ringraziamenti

Ringrazio la mia relatrice, la Professoressa Federica Battisti, di cui ho potuto apprezzare competenza, disponibilità e gentilezza. Ringrazio Elena Camuffo e Luca Hartz per l'aiuto fornitomi nella preparazione dei test. Ringrazio infine la mia famiglia che mi ha sempre sostenuto e supportato.