



**Politecnico
di Torino**

Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria del Cinema e dei Mezzi di Comunicazione

A.a. 2025/2026

Sessione di laurea Luglio 2026

itARTS³

**Sviluppo di una Pipeline Ibrida per lo Storytelling Immersivo
nella Documentazione Audiovisiva del Progetto itARTS**

Relatori:

prof.ssa Tatiana Mazali

prof. Stefano Sburlati

Candidato:

Mattia Saverio Vella

Sommario

Il presente lavoro di tesi si propone di esplorare nuove metodologie di storytelling immersivo applicate all'ambito documentaristico, integrando differenti formati video e tecniche di ricostruzione tridimensionale all'interno di un unico prodotto audiovisivo.

L'elaborato analizza un'intera pipeline produttiva, dall'ideazione alla post-produzione, prendendo in esame il lavoro commissionato a Motion Pixel S.r.l., relativo alla documentazione audiovisiva delle tre principali opere realizzate nell'ambito del progetto nazionale itARTS (Italy for Arts: Resources, Time and Space), finanziato dal PNRR. Nello specifico, l'attività si è concentrata sulla realizzazione di video documentari dedicati alle tre produzioni multidisciplinari nate dalla collaborazione tra l'Accademia di Belle Arti dell'Aquila, il Conservatorio "A. Casella" dell'Aquila e il Politecnico di Torino (*Donne all'Opera*, *Dei Maestri Dipintori* e *Shared*), oltre a un quarto video documentario di presentazione del progetto e di breve sintesi delle tre esperienze.

La metodologia adottata integra riprese stereoscopiche a 180 gradi in 8K 3D, acquisite mediante una videocamera Canon R5 C equipaggiata con obiettivo Dual Fisheye, impiegate per le interviste ai protagonisti, e riprese monoscopiche a 360 gradi, effettuate con due videocamere Qoocam 3 Ultra, al fine di valorizzare l'immersività delle opere teatrali realizzate. Per sfruttare appieno lo spazio visivo circostante durante le scene a 180 gradi e supportare la narrazione, l'ambiente virtuale è stato arricchito con contributi bidimensionali tratti dalle opere stesse quali segni grafici, dipinti, bozzetti di costumi e scenografie, opportunamente integrati nella scena immersiva.

L'esperienza è stata concepita in primis per essere fruita tramite l'utilizzo di un

visore per la realtà virtuale. Al fine di rendere la navigazione tra i contenuti efficace, è stato sviluppato mediante il software 3DVista un contenitore web interattivo che funge da menu di selezione dei video. Tale menu è costruito a partire dalle ricostruzioni tridimensionali dei palcoscenici teatrali associati alle diverse opere, che fanno da sfondo per la selezione dei video.

La ricostruzione 3D degli ambienti del menu è stata realizzata attraverso la tecnica del Gaussian Splatting ed è visualizzata sotto forma di nuvola di punti animata. La creazione dell'ambiente virtuale si basa sull'impiego congiunto del modello neurale Apple SHARP e del modello di stima monoculare della profondità Depth Anything V2, con lo scopo di adattare la pipeline di ricostruzione tridimensionale alle immagini equirettangolari; gli splat ottenuti sono stati successivamente elaborati tramite Blender.

La fase di presentazione e verifica conclusiva del lavoro svolto si è tenuta all'Aquila, in occasione dell'evento conclusivo del progetto itARTS. In tale contesto, i documentari in realtà virtuale sono stati fruiti dal pubblico e dai partecipanti ai laboratori. L'accoglienza è stata ottima e la pipeline tecnica sviluppata ha dimostrato la propria efficacia in funzione dello storytelling documentaristico. Sul piano immersivo, l'integrazione delle diverse tipologie di riprese all'interno di un unico prodotto ha consentito di rafforzare il senso di presenza e di coinvolgimento dello spettatore.

Indice

Elenco delle tabelle	VIII
Elenco delle figure	IX
1 Introduzione	1
1.1 Il Contesto	1
1.1.1 La Realtà Virtuale e la documentazione del Patrimonio Culturale	1
1.1.2 Il Progetto itARTS	3
1.1.3 Il Ruolo Tecnico-Creativo di Motion Pixel S.r.l.	4
1.2 Evoluzione della Documentazione Audiovisiva	5
1.2.1 La relazione spettatore-contenuto: l'evoluzione del concetto di "presence"	5
1.2.2 I formati immersivi: differenze percettive e tecniche tra video 180° e 360°	6
1.3 Obiettivi della Ricerca	7
1.3.1 Lo sviluppo di una "Pipeline Ibrida"	7
1.3.2 Nuovi paradigmi di reality capture: dalla registrazione video alla ricostruzione spaziale della realtà	9
1.4 Struttura dell'Elaborato	9
2 Stato dell'Arte	11
2.1 La Realtà Virtuale nel Patrimonio Culturale e Documentaristico . .	11

2.1.1	La preservazione del patrimonio culturale immateriale: dalla documentazione bidimensionale alla cattura spaziale dello spettacolo dal vivo	11
2.1.2	Il formato 180° stereoscopico nella documentazione d'arte: restituzione tridimensionale del performer e prossimità emotiva	14
2.1.3	Il video 360° monoscopico nello spettacolo dal vivo: documentazione dello spazio scenico e dell'evento performativo globale	15
2.1.4	3D reconstruction: la transizione verso il documentario volumetrico e gli ambienti virtuali navigabili	17
2.2	Tecniche di Acquisizione e Ricostruzione Volumetrica	18
2.2.1	I limiti della fotogrammetria classica e l'avvento dei Neural Radiance Fields (NeRF)	19
2.2.2	Principi fondamentali e vantaggi prestazionali del 3D Gaussian Splatting (3DGS)	21
2.3	L'Intelligenza Artificiale nella ricostruzione tridimensionale	23
2.3.1	Algoritmi AI per la stima monoculare della profondità (Focus: Depth Anything V2)	24
2.3.2	Adattamento per proiezioni equirettangolari: estensione di modelli prospettici a formati immersivi (Apple SHARP e utilizzo delle Cube-map)	25
3	Metodologia e Sviluppo della Pipeline Ibrida	28
3.1	Acquisizione dei dati e setup di ripresa sul set	28
3.1.1	Riprese stereoscopiche	30
3.1.2	Riprese monoscopiche a 360° (Qoocam 3 Ultra)	34
3.2	Post-produzione e organizzazione dei flussi di lavoro	35
3.2.1	Gestione dei dati, stitching e ottimizzazione dei file ad alta risoluzione	35
3.2.2	Workflow collaborativo e archiviazione condivisa tra diverse postazioni di calcolo	35
3.2.3	Integrazione dei contributi bidimensionali delle opere nelle scene a 180°	36

3.2.4	Montaggio e grammatica delle transizioni immersive	37
3.3	Ricostruzione volumetrica e spaziale degli scenari performativi	40
3.3.1	Sperimentazione e confronto delle metodologie di reality capture	41
3.3.2	Stima della profondità (Depth Anything V2 e Apple SHARP) e generazione degli splat gaussiani / nuvola di punti	42
3.3.3	Ottimizzazione e modellazione degli asset tridimensionali in Blender	43
3.4	Architettura e sviluppo dell'interfaccia VR (3DVista)	46
4	Risultati e Conclusioni	49
4.1	I deliverable finali della documentazione immersiva	49
4.2	Discussione dei risultati: sfide tecniche e scelte creative	50
4.3	Validazione empirica all'evento itARTS dell'Aquila	52
4.4	Conclusioni e sviluppi futuri	54
	Bibliografia	55

Elenco delle tabelle

2.1	Confronto sintetico tra le tre famiglie di tecniche di ricostruzione tridimensionale.	24
3.1	Quadro riassuntivo dei sistemi di acquisizione impiegati sul set e dei relativi flussi di pre-elaborazione.	31
4.1	I quattro deliverable del progetto. Ciascun documentario integra le interviste in VR180 stereoscopico e le riprese a 360° di spettacoli e prove; le caratteristiche tecniche di consegna, comuni a tutti, sono riportate nella Tabella 4.2.	50
4.2	Caratteristiche tecniche dei quattro file master, misurate direttamente sui contenitori finali (codifica video HEVC 5760×5760 px, 25 fps; audio AAC stereo, 48 kHz). Le durate sono comprensive dei titoli di coda.	50

Elenco delle figure

1.1	Il <i>Reality-Virtuality Continuum</i> proposto da Milgram e Kishino [1].	2
2.1	L'ottica <i>Canon RF 5.2mm F2.8L Dual Fisheye</i> : a sinistra lo schema degli elementi ottici, con la <i>baseline</i> stereoscopica di 60 mm tra i due assi ottici; a destra l'obiettivo.	14
2.2	Esempio di ripresa frontale e ravvicinata in VR180 stereoscopico: un fotogramma (monoscopico) dal documentario <i>Shared</i>	15
2.3	Esempio di ripresa a 360° che restituisce l'intero spazio scenico: un fotogramma (monoscopico) dal documentario <i>Dei Maestri Dipintori</i>	16
2.4	La pipeline fotogrammetrica classica: dalla stima delle pose (SfM) alla densificazione (MVS) fino alla mesh texturizzata.	19
2.5	Il NeRF come funzione plenottica appresa: una rete MLP mappa posizione e direzione di vista in colore e densità; la sintesi di una nuova vista avviene per <i>volume rendering</i> lungo i raggi (<i>ray marching</i>).	21
2.6	Dal triangolo rasterizzato allo <i>splatting</i> : una singola gaussiana proiettata come impronta ellittica e la composizione di più gaussiane sovrapposte tramite <i>alpha blending</i>	23
2.7	Dalla proiezione equirettangolare alla <i>cube-map</i> : il panorama a 360° dell'Auditorium del Parco (a sinistra) e la sua scomposizione in facce prospettiche (a destra), primo passaggio della pipeline SHARP_360_to_Splat.	27

3.1	Schema d'insieme della pipeline ibrida: il flusso di produzione dei documentari (a sinistra) e il ramo di ricostruzione degli ambienti a nuvola di punti (a destra) convergono nel menu immersivo realizzato in 3DVista. Per chiarezza sono omessi i dispositivi di riserva (B camera e audio di sicurezza).	29
3.2	La troupe al lavoro durante le riprese delle interviste ai protagonisti del progetto.	30
3.3	L'attrezzatura impiegata per le interviste: la Canon EOS R5 C (a sinistra), la DJI Osmo Pocket usata come B camera (al centro) e il registratore ambisonico Zoom H3-VR (a destra).	32
3.4	Insta360 Pro2.	33
3.5	Kandao Qoocam 3 Ultra.	34
3.6	Fotogramma monoscopico di <i>Donne all'Opera</i> : al centro la scena VR180 con l'intervista, mentre nell'emisfero posteriore (ai lati) sono integrati i contributi bidimensionali dell'opera.	37
3.7	La post-produzione in Adobe Premiere Pro: in alto il <i>compositing</i> dei contributi 2D/VR nel documentario (con gli effetti di proiezione immersiva e le grafiche delle arie); in basso il montaggio del video di presentazione.	38
3.8	L'interfaccia della pipeline <i>SHARP_360_to_Splat</i> : il panorama equirettangolare in input, la scomposizione in quattro facce (<i>Sides</i>) e la normalizzazione della scala con DA360 (<i>Alignment Mode</i>), con esportazione dello <i>splat</i> in formato .PLY.	42
3.9	Elaborazione della nuvola in Blender: in alto il risultato nel viewport, con il materiale emissivo azzurro; in basso l'albero di <i>Geometry Nodes</i> (<i>Mesh to Points</i> , <i>Set Position</i> pilotato da una <i>Noise Texture</i> 4D e <i>Set Material</i>).	44
3.10	Il <i>Graph Editor</i> con le curve di animazione dei modificatori (<i>Angle</i> del <i>Simple Deform</i> e <i>Strength</i> del <i>Lattice</i>); il modificatore <i>Cycles</i> rende l'animazione ciclica sui 350 fotogrammi.	45

3.11	Render finale a nuvola di punti di uno degli ambienti ricostruiti (Accademia di Belle Arti dell'Aquila), in proiezione equirettangolare a 360°, con materiale emissivo azzurro e nebbia di profondità. . . .	45
3.12	Ambiente di authoring in 3DVista: il menu immersivo con lo sfondo animato a nuvola di punti a 360° e gli <i>hotspot</i> di selezione dei quattro documentari.	47
4.1	Il formato di consegna: fotogramma <i>Top-and-Bottom</i> (occhio sinistro sopra, destro sotto), 5760×5760 pixel (<i>Dei Maestri Dipintori</i>). . . .	49
4.2	Spettatori eterogenei provano i documentari immersivi con il visore Meta Quest 3 durante l'evento conclusivo del progetto itARTS all'Aquila.	53

Capitolo 1

Introduzione

Negli ultimi anni i media immersivi hanno progressivamente ridefinito il modo in cui le esperienze visive vengono prodotte e fruita, spostando il baricentro della comunicazione audiovisiva dallo schermo bidimensionale allo spazio che circonda lo spettatore. La possibilità di collocare l'osservatore al centro della scena, e non più di fronte ad essa, ha aperto prospettive inedite non soltanto per l'intrattenimento, ma anche per la documentazione e la conservazione del patrimonio culturale, in particolare di quelle forme d'arte performative la cui natura effimera ne ha storicamente reso difficile ogni tentativo di archiviazione. È in questo scenario in rapida evoluzione che si inserisce il presente lavoro di tesi, il quale parte da una riflessione generale sui linguaggi immersivi per giungere successivamente all'analisi di una specifica esperienza produttiva, sviluppata nell'ambito del progetto itARTS. Prima di entrare nel merito del caso di studio, è opportuno delineare il contesto tecnologico e culturale entro cui esso ha preso forma.

1.1 Il Contesto

1.1.1 La Realtà Virtuale e la documentazione del Patrimonio Culturale

Nel corso degli ultimi decenni l'evoluzione dei sistemi di calcolo e delle tecnologie di visualizzazione ha progressivamente eroso i confini bidimensionali dello schermo,

ridefinendo il rapporto tra l'osservatore e l'immagine. Un riferimento concettuale tuttora imprescindibile per inquadrare questa trasformazione è il *Reality-Virtuality Continuum* proposto da Milgram e Kishino nel 1994, che descrive uno spettro continuo compreso tra l'ambiente reale e quello interamente sintetico, all'interno del quale si collocano le diverse forme di Realtà Mista (Mixed Reality), Aumentata (Augmented Reality) e Virtualità Aumentata (Augmented Virtuality) [1]. All'estremo virtuale di tale continuum si situa la Realtà Virtuale (VR), intesa come tecnologia immersiva capace di sostituire integralmente gli stimoli visivi e spaziali dell'ambiente fisico con quelli di un ambiente simulato ed esplorabile, nel quale l'utente non osserva più la scena dall'esterno ma ne diviene parte integrante.

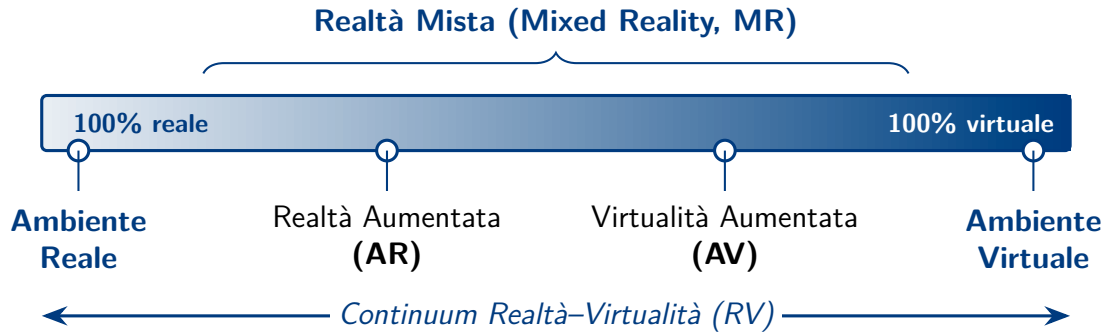


Figura 1.1: Il *Reality-Virtuality Continuum* proposto da Milgram e Kishino [1].

Applicata al dominio dei beni culturali, la Realtà Virtuale ha reso possibili nuovi paradigmi nelle pratiche di documentazione e conservazione: si è passati dai primi sistemi *low-immersive*, fruiti su schermi convenzionali e caratterizzati da un'interazione limitata, agli attuali ambienti *high-immersive* basati su visori (Head-Mounted Display), che consentono la ricostruzione e l'esplorazione attiva di luoghi e contesti del patrimonio [2]. La diffusione di dispositivi di visualizzazione di tipo consumer e di motori di sviluppo come Unity e Unreal Engine ha reso possibile il passaggio da una documentazione di stampo archivistico (lineare, passiva e bidimensionale) a una forma di conservazione interattiva e partecipativa, in cui la qualità dell'esperienza si misura sul senso di presenza e di coinvolgimento dell'utente [2, 3].

Tale transizione assume un rilievo particolare nel caso del patrimonio culturale immateriale. La Convenzione UNESCO per la salvaguardia del patrimonio culturale

immateriale del 2003 riconosce esplicitamente tra i propri domini le arti dello spettacolo e definisce la salvaguardia come l'insieme delle misure volte a garantirne la vitalità, ivi compresa la documentazione [4]. Diversamente dal patrimonio materiale, le espressioni performative (teatro, musica, danza) possiedono una natura intrinsecamente effimera: esistono solo nell'atto della loro esecuzione e si dissolvono al suo termine, rendendo storicamente problematico ogni tentativo di registrazione e trasmissione. La letteratura sulla digitalizzazione del patrimonio immateriale sottolinea inoltre come una documentazione efficace non possa limitarsi a isolare il gesto del performer o l'opera, ma debba restituire l'intero contesto spaziale e relazionale in cui la pratica culturale si iscrive [5].

È in questa prospettiva che si collocano le tecniche di *reality capture*. Metodologie ormai consolidate, come la fotogrammetria e la scansione laser tridimensionale, da tempo impiegate nel rilievo e nella modellazione di siti e manufatti del patrimonio culturale [6], hanno reso possibile l'acquisizione e la conservazione dell'intera volumetria spaziale degli oggetti analizzati. L'integrazione di dati geometrici tridimensionali all'interno di ambienti immersivi navigabili consente così di superare i limiti della rappresentazione piana, restituendo all'osservatore la libertà di esplorare la scena e un senso di presenza psicologica difficilmente raggiungibile attraverso i media tradizionali [2].

Ne deriva una possibilità inedita per le arti performative, che trovano una nuova modalità di conservazione e fruizione. Questo approccio apre frontiere significative per la divulgazione e la tutela del patrimonio culturale immateriale italiano e costituisce la premessa concettuale entro cui si inserisce il lavoro documentato nel presente elaborato.

1.1.2 Il Progetto itARTS

Il lavoro documentato nel presente elaborato si inserisce nel quadro del progetto nazionale **itARTS: Italy for Arts: Resources, Time and Space**, finanziato dall'Unione Europea nell'ambito del Piano Nazionale di Ripresa e Resilienza, dedicato all'innovazione della dimensione internazionale del sistema dell'Alta Formazione Artistica, Musicale e Coreutica (AFAM). Coordinato dal Conservatorio di Musica "Alfredo Casella" dell'Aquila in qualità di capofila, il progetto riunisce un consorzio

di istituzioni accademiche: l'Accademia di Belle Arti dell'Aquila, l'Università degli Studi dell'Aquila, il Saint Louis College of Music di Roma e il Politecnico di Torino, attraverso il corso di Laurea in Ingegneria del Cinema e dei Mezzi di Comunicazione, con l'obiettivo di valorizzare il patrimonio culturale italiano e la ricerca artistica nella pratica performativa.

Nell'ambito di questo consorzio, la presente tesi si concentra sulla collaborazione tra tre istituzioni: il Politecnico di Torino, l'Accademia di Belle Arti dell'Aquila e il Conservatorio "A. Casella". Tale collaborazione ha dato vita a tre distinte produzioni multidisciplinari:

Donne all'Opera (Croce e Delizia), dedicata alla reinterpretazione e alla valorizzazione delle figure femminili nel contesto operistico e teatrale;

Dei Maestri Dipintori, incentrata sul dialogo visivo e performativo ispirato alla grande tradizione pittorica;

Shared, orientata alla co-creazione e all'interazione tra pubblico e performer, in cui i gesti vengono trasformati in suono e forme visive; il lavoro si articola in una componente fisica e in una declinazione in Realtà Virtuale (*SharedVR*) sviluppata dal Politecnico di Torino.

L'acronimo **itARTS**³, utilizzato nel titolo della presente tesi, nasce proprio per sintetizzare questa architettura progettuale: l'esponente al cubo rappresenta infatti sia l'unione delle *tre* istituzioni accademiche coinvolte (Politecnico di Torino, Accademia di Belle Arti dell'Aquila, Conservatorio "A. Casella"), sia le *tre* distinte opere performative appena citate.

All'interno di questo scenario, l'obiettivo centrale del presente studio è stato lo sviluppo di quattro documentari immersivi finalizzati alla narrazione delle opere e dell'iniziativa nel suo complesso. Per la produzione di questo materiale audiovisivo, ci si è avvalsi della società Motion Pixel S.r.l., specializzata nella produzione di contenuti in Realtà Virtuale, il cui ruolo tecnico-creativo è approfondito nel paragrafo seguente.

1.1.3 Il Ruolo Tecnico-Creativo di Motion Pixel S.r.l.

La realizzazione della documentazione audiovisiva delle opere itARTS è stata affidata a **Motion Pixel S.r.l.**, società specializzata nella produzione di contenuti

immersivi in Realtà Virtuale, guidata dal professor Stefano Sburlati, co-relatore del presente lavoro di tesi.

La sfida principale ha riguardato il superamento della documentazione audiovisiva tradizionale e lineare, basata su formati bidimensionali come il 16:9 orizzontale o i più recenti tagli verticali dei contenuti social.

L'obiettivo tecnico-creativo è stato la progettazione e lo sviluppo di un'esperienza immersiva in Realtà Virtuale capace di documentare non soltanto la messinscena finale, ma anche il processo creativo dietro ogni opera.

Il progetto di tesi si è focalizzato sull'intera filiera produttiva, dall'ideazione fino alla presentazione finale. La fase di pre-produzione è stata condotta insieme al prof. Sburlati; le riprese si sono svolte nell'arco di più giornate all'Aquila e ad Annecy, affidate di volta in volta a un ristretto gruppo di operatori che ha visto la collaborazione, in momenti differenti, anche di Matteo Pellegrini e del professor Mattia Meloni; la post-produzione è stata curata principalmente dal sottoscritto insieme al prof. Sburlati, con il supporto del prof. Meloni per la color correction. Questo coinvolgimento diretto in ogni fase del processo costituisce il fondamento esperienziale su cui si basa l'analisi della pipeline produttiva sviluppata nei capitoli seguenti.

1.2 Evoluzione della Documentazione Audiovisiva

1.2.1 La relazione spettatore-contenuto: l'evoluzione del concetto di “presence”

L'affermazione dei media immersivi introduce differenze significative rispetto alla documentazione audiovisiva bidimensionale: il controllo che il regista esercita sull'inquadratura si riduce, sostituito da forme più indirette di orchestrazione dell'attenzione. In questo contesto, la misura che meglio descrive l'efficacia dell'esperienza non è più la qualità della singola inquadratura, bensì il senso di presenza (*presence*), inteso come la sensazione psicologica di “essere lì”, all'interno dell'ambiente rappresentato.

La presenza (*presence*) costituisce la risposta soggettiva e percettiva generata da un'esperienza immersiva, manifestandosi come la sensazione psicologica di trovarsi realmente all'interno dell'ambiente rappresentato. Diversamente dall'immersione (*immersion*), che descrive caratteristiche oggettive del sistema tecnologico, quali campo visivo, risoluzione, latenza e stimolazione multisensoriale, la presenza dipende anche dal contenuto e dalle modalità con cui l'esperienza viene progettata [7].

All'interno di questo quadro teorico, il modello proposto da Mel Slater scompone l'esperienza di presenza in due illusioni complementari e ortogonali [8]. La prima, definita *Place Illusion* (PI), corrisponde all'illusione di trovarsi effettivamente nel luogo virtuale e si attiva quando il sistema risponde in modo coerente alle contingenze senso-motorie dell'utente, restituendo, ad esempio, la porzione corretta di scena al variare della posizione o dell'orientamento del capo. La seconda, la *Plausibility Illusion* (Psi), riguarda invece l'illusione che gli eventi rappresentati stiano realmente accadendo e dipende dalla coerenza interna, dalla credibilità e dalla reattività dell'ambiente virtuale. Quando entrambe le illusioni risultano sufficientemente sostenute, l'utente tende a reagire allo spazio virtuale in modo analogo a quanto farebbe in un contesto reale [8].

Nel contesto della documentazione delle arti performative, questo significa che la qualità dell'esperienza non dipende soltanto dalla precisione tecnica della cattura, ma soprattutto dalla capacità di restituire le condizioni percettive e relazionali proprie della fruizione dal vivo. L'obiettivo, quindi, non è semplicemente registrare l'evento scenico, ma ricreare una relazione credibile tra spettatore, performer e spazio, così da trasformare la visione passiva di una registrazione in un'esperienza immersiva e coinvolgente. È a partire da questo principio che sono state orientate le scelte di formato, montaggio e interazione descritte nel presente elaborato.

1.2.2 I formati immersivi: differenze percettive e tecniche tra video 180° e 360°

La realizzazione concreta della *Place Illusion* passa attraverso la scelta della tecnologia di ripresa. Il formato a 360 gradi offre il vantaggio della completa libertà di esplorazione visiva: l'utente può orientare lo sguardo in qualsiasi direzione, condizione preziosa per documentare l'architettura scenica e la dimensione corale

di una performance. Questa stessa ampiezza comporta però alcuni limiti. Sul piano percettivo, distribuendo un numero finito di pixel sull'intera superficie sferica, il formato riduce la densità angolare dell'immagine misurata in pixel per grado (PPD) con un conseguente calo di nitidezza rispetto a una ripresa che copra un campo più ristretto; inoltre, quando l'acquisizione è monoscopica, viene a mancare del tutto la parallasse stereoscopica e con essa la percezione della profondità. Sul piano tecnico, inoltre, l'acquisizione sferica richiede la cucitura (*stitching*) dei flussi provenienti da più obiettivi, operazione che introduce artefatti nelle zone di sovrapposizione e in prossimità dei poli; tali criticità risultano ancora più marcate nelle configurazioni stereoscopiche, nelle quali è necessario mantenere una corrispondenza geometrica tra le due viste lungo l'intera sfera.

Il formato VR180 segue una logica complementare: concentra l'intera risoluzione del sensore nell'emisfero anteriore, quello effettivamente osservato da un utente seduto, e cattura la scena attraverso due obiettivi distanziati in modo da imitare la distanza interpupillare umana, restituendo così una profondità stereoscopica nativa. Poiché il campo visivo binoculare umano, responsabile della percezione della profondità, si estende per circa 114 gradi, una copertura di 180 gradi è più che sufficiente a saturare l'area di visione utile, eliminando al contempo gli artefatti di cucitura posteriori tipici del 360. Ne deriva un formato particolarmente adatto alle riprese ravvicinate e frontali (le interviste ai protagonisti, in primo luogo) dove la tridimensionalità e la prossimità contribuiscono a una forte intimità percettiva tra soggetto e spettatore.

I due formati rispondono dunque a esigenze narrative differenti e in larga parte complementari: è da questa complementarità che muove la pipeline ibrida proposta in questo lavoro.

1.3 Obiettivi della Ricerca

1.3.1 Lo sviluppo di una “Pipeline Ibrida”

L'obiettivo principale della presente ricerca consiste nello sviluppo e nell'applicazione di una pipeline ibrida di ripresa e post-produzione per la realizzazione di documentazione audiovisiva immersiva in Realtà Virtuale. Più che rappresentare

un esito teorico, la pipeline costituisce lo strumento operativo attraverso cui dare forma concreta all'esperienza documentaria. Il suo carattere ibrido risponde a una necessità emersa nel corso del lavoro: un singolo formato risulta difficilmente sufficiente a restituire la complessità dell'evento performativo, nella sua duplice dimensione intima e spaziale.

A questa motivazione di ordine narrativo e percettivo si affiancano inoltre ragioni di natura pratico-produttiva: la varietà dei supporti impiegati risponde anche a esigenze di maneggevolezza, rapidità di setup e discrezione sul set, aspetti che verranno approfonditi nel Capitolo 3.

L'approccio adottato ha quindi cercato di assegnare a ciascuna tecnologia il compito per cui risulta più efficace, senza tuttavia imporre una divisione rigida. Si è data priorità all'impiego delle riprese a 180 gradi stereoscopiche per i momenti a più alto carico emotivo: le interviste ai protagonisti, gli interventi dei docenti e dei registi, dove la profondità e la frontalità rafforzano la connessione con lo spettatore, riservando alle riprese a 360 gradi la restituzione dello spazio scenico e l'esplorazione del contesto performativo. Questa ripartizione, però, non è stata applicata in maniera netta: in alcune scene di ambientazione il formato a 180 gradi si è rivelato più efficace ed è stato perciò impiegato anche al di fuori delle interviste. A questi contributi video si affiancano l'integrazione di materiali bidimensionali tratti dalle opere e, per il menu di fruizione, la ricostruzione tridimensionale degli ambienti.

Combinare formati diversi all'interno della stessa esperienza comporta una serie di complessità tecniche, soprattutto in fase di post-produzione. Differenze di proiezione, risoluzione e linguaggio visivo devono essere gestite con attenzione, in particolare nel montaggio, dove transizioni troppo brusche possono compromettere il comfort percettivo dello spettatore e generare affaticamento o chinetosi (*motion sickness*). La messa a punto di una pipeline capace di gestire tali criticità rappresenta uno dei principali contributi di questa ricerca.

1.3.2 Nuovi paradigmi di reality capture: dalla registrazione video alla ricostruzione spaziale della realtà

Il secondo obiettivo della ricerca riguarda l'esplorazione di tecniche di ricostruzione tridimensionale degli spazi teatrali, utilizzati come sfondo animato del menu dell'applicazione. Quest'ultimo è stato progettato per consentire all'utente di selezionare i diversi documentari all'interno di un'unica esperienza. In quest'ottica, i palcoscenici sono stati ricreati come ambienti immersivi dall'estetica a nuvola di punti, in cui i punti si aggregano e si dissolvono progressivamente per dare forma ai diversi teatri e costruire transizioni fluide tra l'uno e l'altro. Le ricostruzioni sono state renderizzate come video monoscopici a 360 gradi, mantenendo la videocamera virtuale fissa per evitare fenomeni di chinetosi e affidando il movimento esclusivamente all'animazione della nuvola di punti, in continuità con i tre gradi di libertà (3-DoF) adottati nei documentari.

La generazione di questi ambienti si è avvalsa dei più recenti modelli neurali di computer vision: la stima monoculare della profondità di Depth Anything V2 [9] e la sintesi tridimensionale a partire da una singola immagine di SHARP, modello sviluppato da Apple [10]. A differenza delle tecniche di ricostruzione consolidate, la fotogrammetria, i Neural Radiance Fields (NeRF) e lo stesso 3D Gaussian Splatting [11], sperimentate anch'esse nel corso del lavoro, che richiedono numerose immagini riprese da punti di vista differenti e tempi di elaborazione nell'ordine delle ore, questi modelli consentono di ottenere un risultato qualitativamente paragonabile a partire da una sola immagine equirettangolare, generando ciascuno *splat* nell'ordine dei secondi. Gli ambienti così ricostruiti sono stati infine rifiniti in Blender e integrati come sfondi animati nell'ambiente di authoring 3DVista; il loro inquadramento nello stato dell'arte e il dettaglio della pipeline sono approfonditi rispettivamente nei Capitoli 2 e 3.

1.4 Struttura dell'Elaborato

L'elaborato è organizzato in quattro capitoli. Il presente Capitolo 1 ha introdotto il contesto culturale e tecnologico della ricerca, il progetto itARTS e gli obiettivi del lavoro. Il Capitolo 2 ricostruisce lo stato dell'arte scientifico e tecnologico,

esaminando l'evoluzione del documentario in Realtà Virtuale e dello storytelling immersivo, le tecniche di ricostruzione volumetrica (dalla fotogrammetria ai Neural Radiance Fields fino al 3D Gaussian Splatting) e i modelli di intelligenza artificiale per la stima della profondità e la sintesi tridimensionale, con particolare attenzione al loro adattamento alle proiezioni equirettangolari. Il Capitolo 3 descrive in dettaglio la pipeline ibrida sviluppata: l'acquisizione dei dati sul set, i flussi di post-produzione, la ricostruzione volumetrica degli scenari e lo sviluppo dell'interfaccia VR in 3DVista. Il Capitolo 4 presenta infine i deliverable finali, la loro presentazione e validazione in occasione dell'evento conclusivo del progetto itARTS a L'Aquila, una discussione critica dei punti di forza e dei limiti della pipeline e le prospettive di sviluppo futuro.

Capitolo 2

Stato dell'Arte

Questo capitolo traccia il quadro teorico e tecnologico su cui poggia la pipeline ibrida presentata nei capitoli seguenti, partendo dalla documentazione del patrimonio performativo in Realtà Virtuale e dai formati immersivi impiegati per realizzarla, per poi passare alle tecniche di acquisizione e ricostruzione volumetrica della scena, dalla fotogrammetria fino al 3D Gaussian Splatting. Infine verranno trattati i modelli di intelligenza artificiale per la stima della profondità e la sintesi tridimensionale con particolare attenzione al loro adattamento alle immagini equirettangolari, fino alla soluzione concretamente adottata in questo lavoro.

2.1 La Realtà Virtuale nel Patrimonio Culturale e Documentaristico

2.1.1 La preservazione del patrimonio culturale immateriale: dalla documentazione bidimensionale alla cattura spaziale dello spettacolo dal vivo

La documentazione delle arti performative appartiene al patrimonio culturale immateriale (*Intangible Cultural Heritage*, ICH), caratterizzato da una natura effimera che lo lega indissolubilmente all'atto della sua esecuzione. La rassegna sullo stato dell'arte della digitalizzazione “incarnata” dell'ICH proposta da Hou,

Kenderdine e colleghi evidenzia i limiti degli approcci centrati esclusivamente sulla cattura del gesto del performer attraverso tecniche di motion capture e pose estimation, sottolineando la crescente necessità di forme di documentazione capaci di restituire anche lo spazio, il contesto e le relazioni entro cui la pratica culturale prende forma [5]. La sola cattura del corpo, isolato dal proprio ambiente, si rivela infatti insufficiente a preservare il significato culturale della performance.

È in questa prospettiva che la Realtà Virtuale si è affermata come medium privilegiato. Lo studio di Argyriou, Economou e Bouki, che combina valutazione soggettiva tramite questionari e misurazione oggettiva dell'attività cerebrale mediante elettroencefalografia (EEG), documenta come l'integrazione di esperienze interattive in VR con la narrazione a 360° produca elevati livelli di presenza, immersione e coinvolgimento nella fruizione del patrimonio culturale [12]. Già in precedenza, nel campo dell'*immersive journalism*, de la Peña e colleghi avevano mostrato come la Realtà Virtuale, attraverso la ricostruzione virtuale di eventi reali a partire da materiali documentari, potesse offrire all'utente un'esperienza in prima persona di fatti e testimonianze, fondando esplicitamente tale potenziale documentaristico sul senso di presenza [13]. La realizzazione emblematica di questa stagione pionieristica è *Clouds Over Sidra* (2015) di Gabo Arora e Barry Pousman, primo documentario in Realtà Virtuale prodotto per le Nazioni Unite: il ritratto a 360° della quotidianità di una dodicenne siriana nel campo profughi di Za'atari, presentato al World Economic Forum di Davos, ha dimostrato su scala globale la capacità del mezzo di generare vicinanza empatica verso realtà altrimenti distanti [27]. Nel contesto italiano, un esempio significativo dello stesso potenziale è VR Free (2019) di Milad Tangshir, cortometraggio in Realtà Virtuale a 360° girato all'interno della Casa Circondariale Lorusso e Cutugno di Torino, in cui i detenuti osservano tramite visore scene di vita esterna al carcere. Selezionata alla 76^a Mostra Internazionale d'Arte Cinematografica di Venezia e al Sundance Film Festival, l'opera mostra come la fruizione immersiva possa generare un forte senso di presenza e di prossimità empatica con un contesto reale, confermando il valore del VR come strumento di documentazione e testimonianza [14].

Parallelamente al filone documentaristico, anche il mondo dello spettacolo dal vivo ha iniziato presto a sperimentare la cattura immersiva delle proprie

produzioni. Già nel 2015 Felix & Paul Studios, in coproduzione con Cirque du Soleil, realizzava *Inside the Box of Kurios*, esperienza in 360° stereoscopico premiata con un Daytime Emmy Award, che colloca lo spettatore direttamente sul palcoscenico dello spettacolo, in mezzo ai performer [28]; l'anno successivo il Dutch National Ballet produceva *Night Fall*, primo balletto al mondo concepito e coreografato espressamente per la ripresa a 360°, con l'osservatore posto al centro della scena tra i danzatori [29]. L'esperienza maturata in queste produzioni pionieristiche ha generato anche una prima riflessione autoriale sulla regia immersiva, documentata in interviste e testimonianze dirette dei loro creatori. André Lauzon, produttore di Cirque du Soleil Média e responsabile del progetto *Kurios*, ha ricondotto la gestione dell'attenzione nello spazio sferico non alla grammatica cinematografica, bensì all'esperienza teatrale trentennale della compagnia: come sulla scena dal vivo, ogni momento prevede un soggetto primario e uno secondario, e la composizione alterna deliberatamente sequenze di intensità e di quiete, poiché una stimolazione costante finirebbe per risultare monotona [37]. Sulla stessa linea, Félix Lajeunesse e Paul Raphaël hanno sintetizzato la propria poetica in una formula divenuta emblematica, “*the camera is not a camera, it's a viewer*”, da cui discendono regole operative precise: la macchina da presa resta ferma, perché muoverla equivale a muovere lo spettatore contro la sua volontà; ogni taglio va giustificato, perché ciò che si sta coltivando è la presenza; e l'obiettivo non è mostrare qualcosa, ma generare “*a state of being somewhere instead of watching something*” [38]. Queste posizioni, nate dalla pratica produttiva prima che dalla ricerca accademica, anticipano molti dei principi oggi consolidati nella letteratura sulla regia immersiva (§2.1.3) e trovano riscontro diretto nelle scelte operative adottate nel presente lavoro (§3.2.4).

La letteratura più recente documenta del resto un impiego crescente della Realtà Virtuale nella pratica teatrale contemporanea, tanto nella creazione quanto nella documentazione degli spettacoli [30]. Pur differenti per finalità e linguaggio, queste esperienze condividono con il presente lavoro l'intuizione di fondo: portare lo spettatore all'interno dello spazio performativo, in una posizione preclusa alla fruizione tradizionale in sala. Gli sviluppi più recenti di questa traiettoria, orientati alla cattura congiunta di performance e ambiente mediante ricostruzione volumetrica, sono approfonditi nel §2.1.4.

2.1.2 Il formato 180° stereoscopico nella documentazione d'arte: restituzione tridimensionale del performer e prossimità emotiva

Il formato stereoscopico a 180°, noto come *VR180* e promosso da Google in collaborazione con YouTube nel 2017 [31], cattura il solo emisfero anteriore della scena attraverso due obiettivi fisheye, tipicamente separati da una distanza prossima a quella interpupillare umana, così da ottenere una stereoscopia nativa (Figura 2.1). Rispetto al video sferico a 360°, il VR180 non rappresenta un formato intrinsecamente superiore, ma un diverso compromesso tra copertura spaziale, qualità percettiva e controllo autoriale. Limitando la cattura al solo campo frontale, il formato sacrifica la possibilità di osservare l'intera scena in favore di una maggiore densità di dettaglio nell'area effettivamente fruita dallo spettatore, riducendo al contempo le problematiche di stitching tipiche delle configurazioni sferiche complete. La separazione tra le due viste consente inoltre di restituire indizi binoculari di profondità, particolarmente efficaci nella percezione di soggetti collocati a breve e media distanza.

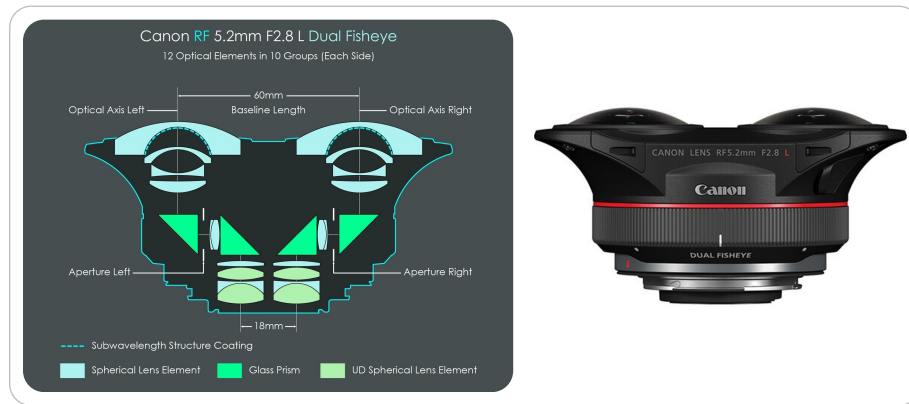


Figura 2.1: L'ottica *Canon RF 5.2mm F2.8L Dual Fisheye*: a sinistra lo schema degli elementi ottici, con la *baseline* stereoscopica di 60 mm tra i due assi ottici; a destra l'obiettivo.

Per queste caratteristiche, il VR180 risulta particolarmente adatto a forme di narrazione frontale e ravvicinata, come interviste, dialoghi e ritratti (Figura 2.2), in cui prossimità e stereoscopia contribuiscono a rafforzare il senso di presenza e la relazione percettiva tra soggetto e spettatore.

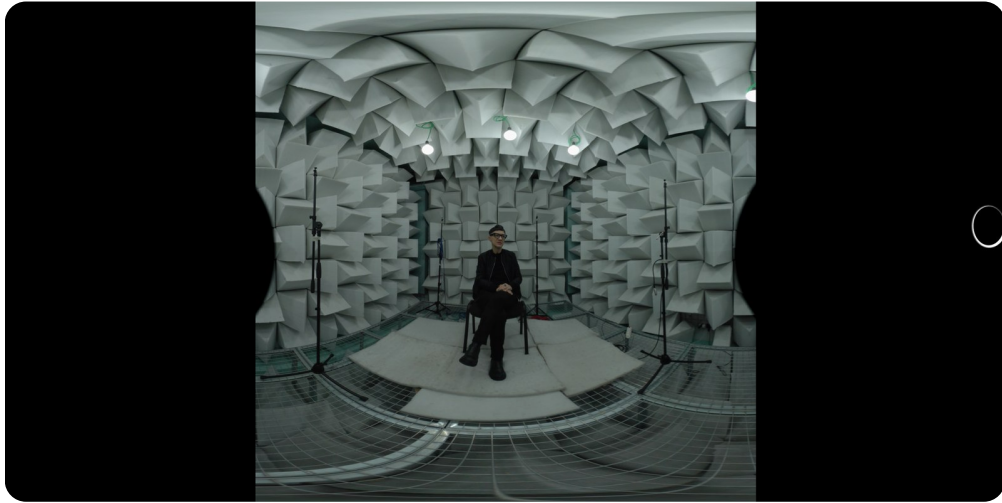


Figura 2.2: Esempio di ripresa frontale e ravvicinata in VR180 stereoscopico: un fotogramma (monoscopico) dal documentario *Shared*.

Rispetto al video sferico, il formato offre inoltre condizioni più favorevoli al controllo della composizione visiva, poiché concentra l'attenzione su un campo d'azione delimitato, riducendo la dispersione dello sguardo. La crescente diffusione di sistemi di ripresa dedicati e il rinnovato interesse dell'industria verso contenuti immersivi stereoscopici, culminato nell'introduzione del formato *Apple Immersive Video* (video stereoscopico a 180° in 8K con audio spaziale, destinato al visore Apple Vision Pro) [32], confermano la rilevanza del VR180 come formato intermedio tra la grammatica cinematografica tradizionale e l'esperienza pienamente sferica della Realtà Virtuale.

2.1.3 Il video 360° monoscopico nello spettacolo dal vivo: documentazione dello spazio scenico e dell'evento performativo globale

Il video sferico a 360° distribuisce l'informazione visiva sull'intero angolo solido, consentendo allo spettatore di esplorare liberamente l'ambiente. Questa caratteristica lo rende particolarmente adatto alla documentazione dell'architettura scenica e della dimensione corale di una performance, permettendo di restituire una percezione immersiva dello spazio teatrale nella sua interezza [12] (Figura 2.3).

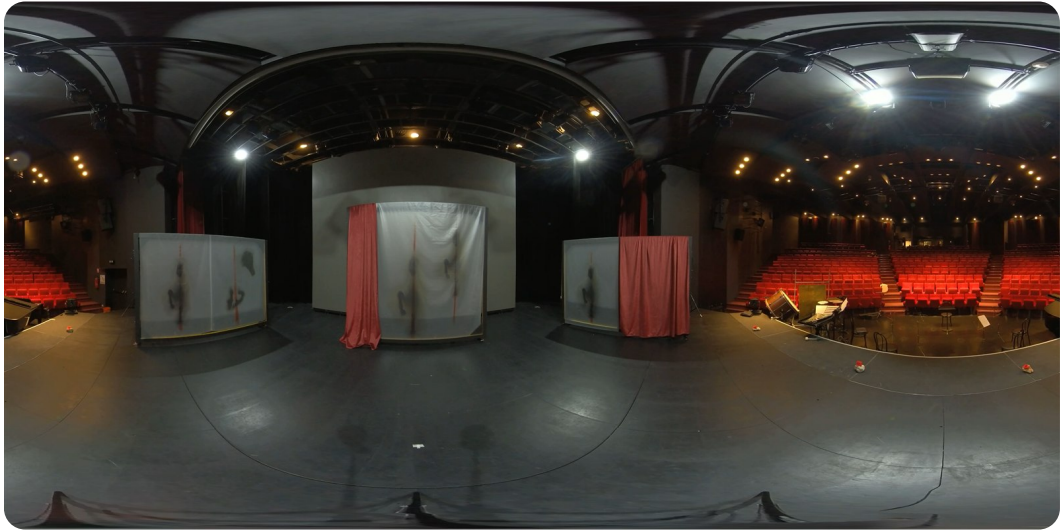


Figura 2.3: Esempio di ripresa a 360° che restituisce l'intero spazio scenico: un fotogramma (monoscopico) dal documentario *Dei Maestri Dipintori*.

Il formato presenta tuttavia limiti ben documentati in letteratura, di natura narrativa, percettiva e tecnica. Sul piano narrativo, l'assenza di una cornice visiva fissa e quindi di un vero e proprio fuori campo riduce il controllo autoriale sulla direzione dello sguardo, imponendo al regista di ripensare gli strumenti tradizionali del proprio mestiere [33], con il rischio che eventi rilevanti si svolgano al di fuori del campo visivo istantaneo dell'utente. Diversi studi sulla regia immersiva mostrano inoltre come gli spettatori tendano a esplorare solo parzialmente la sfera, mantenendo spesso un orientamento prevalentemente frontale, in particolare nella fruizione da seduti. Ciò rende necessarie strategie esplicite di guida dell'attenzione, come l'impiego del suono spazializzato, il movimento degli elementi in scena o tecniche di *misdirection*, capaci di orientare lo spettatore verso i punti salienti senza comprometterne la libertà esplorativa [16, 34]. La libertà di osservazione rappresenta quindi, al tempo stesso, il principale punto di forza e una delle maggiori sfide registiche del formato.

Sul piano percettivo, la variante monoscopica non fornisce indizi binoculari di profondità e restituisce pertanto una percezione spaziale meno efficace rispetto ai formati stereoscopici. Inoltre, la distribuzione di un numero finito di pixel sull'intera sfera riduce drasticamente la densità angolare dell'immagine (*pixels per degree*, PPD) rispetto a un'inquadratura tradizionale, imponendo risoluzioni

di acquisizione molto elevate (tipicamente nell'ordine dell'8K o superiori) per garantire una nitidezza adeguata alla visione in headset. Sul piano tecnico, infine, l'acquisizione sferica richiede la cucitura (*stitching*) dei flussi provenienti da più obiettivi in un'unica immagine equirettangolare. Poiché gli obiettivi occupano posizioni fisiche leggermente diverse, uno stesso oggetto viene proiettato in modo non coincidente nelle rispettive aree di sovrapposizione: questa disparità di parallasse, tanto più marcata quanto più il soggetto è vicino alla camera, si traduce in disallineamenti, sdoppiamenti (*ghosting*) e discontinuità visibili lungo le linee di giunzione, particolarmente critiche in prossimità dei poli. Il problema si acuisce nelle configurazioni stereoscopiche, che devono preservare la corrispondenza geometrica tra le due viste sull'intera sfera.

2.1.4 3D reconstruction: la transizione verso il documentario volumetrico e gli ambienti virtuali navigabili

Accanto al video sferico, si sono sviluppati approcci di acquisizione orientati alla ricostruzione geometrica tridimensionale della scena. A differenza del video a 360°, che pur avvolgendo lo spettatore rimane vincolato a un unico punto di vista e consente una fruizione limitata a tre gradi di libertà (3-DoF) corrispondenti ai movimenti rotatori del capo lungo gli assi di *yaw*, *pitch* e *roll*, la ricostruzione tridimensionale dell'ambiente apre, almeno in linea teorica, alla possibilità di esperienze volumetriche esplorabili a sei gradi di libertà (6-DoF). Ai tre movimenti rotatori si aggiungono infatti tre traslazioni nello spazio (avanti/indietro, sinistra/destra e alto/basso) che permettono all'osservatore non soltanto di orientare lo sguardo, ma anche di spostarsi fisicamente all'interno della scena.

La cattura di scene dinamiche esplorabili a sei gradi di libertà è oggetto di una linea di ricerca attiva, che spazia dai sistemi di *volumetric video* basati su array di telecamere calibrate [35] fino alle recenti estensioni dinamiche del *Gaussian Splatting*, come il *4D Gaussian Splatting*, che introduce la dimensione temporale nella rappresentazione per rendere in tempo reale scene in movimento [36]. Tali approcci rappresentano la frontiera verso un documentario performativo pienamente volumetrico; restano tuttavia onerosi (richiedono numerose viste sincronizzate e ingenti risorse di calcolo) e ancora imperfetti nella resa di soggetti umani in

movimento. È anche in ragione di questi limiti che il presente lavoro adotta una soluzione ibrida e pragmatica, affidando la performance al video immersivo e riservando la ricostruzione tridimensionale ai soli ambienti statici.

In ambito performativo, il precedente più prossimo al presente lavoro è il paradigma di *living documentation* proposto da Tian e colleghi [15], che integra la performance della Kunqu (opera tradizionale cinese) con l'ambiente tangibile dei giardini classici di Suzhou in un "teatro digitale integrato". Attraverso array multi-camera e *Gaussian Splatting* (combinati con video, motion capture e modelli 3D), gli autori catturano congiuntamente il gesto performativo e il suo contesto architettonico, mostrando in uno studio con gli utenti come la restituzione integrata migliori spazialità, realismo e coerenza culturale rispetto alle ricostruzioni basate su fotogrammetria e motion capture tradizionali.

Il presente lavoro condivide con questo paradigma il rifiuto di isolare il performer dal proprio ambiente, ma se ne discosta nella strategia di cattura. Laddove la *living documentation* applica il *Gaussian Splatting* alla performance stessa (con i costi, i requisiti multi-camera e gli artefatti che la ricostruzione volumetrica di soggetti umani in movimento ancora comporta), la pipeline qui proposta mantiene la performance nel dominio del video immersivo stereoscopico, che ne preserva fedeltà, volti e dinamica, e riserva la ricostruzione a nuvola di punti ai soli spazi teatrali statici, impiegandola come sfondo evocativo del menu di fruizione. Ne risulta una divisione del lavoro deliberata: il video dove contano fedeltà e movimento, il *Gaussian Splatting* dove conta la restituzione atmosferica dello spazio. Le tecniche che rendono possibile tale ricostruzione (dalla fotogrammetria ai *neural radiance fields*) sono approfondite nel §2.2.

2.2 Tecniche di Acquisizione e Ricostruzione Volumetrica

La realizzazione di ambienti tridimensionali a partire da immagini ha conosciuto negli ultimi anni una rapida evoluzione, passando dalle tecniche fotogrammetriche tradizionali ai metodi di *neural rendering*. Questa sezione ne ripercorre la traiettoria,

dai limiti della fotogrammetria fino al 3D Gaussian Splatting, la tecnica su cui si fonda la ricostruzione degli ambienti impiegati in questo lavoro.

2.2.1 I limiti della fotogrammetria classica e l'avvento dei Neural Radiance Fields (NeRF)

La fotogrammetria rappresenta da decenni uno standard consolidato per il rilievo tridimensionale metrico del patrimonio culturale [6]. Il processo si articola in due fasi principali. Nella prima, la *Structure-from-Motion* (SfM) ricostruisce simultaneamente la geometria sparsa della scena e la posizione e l'orientamento (*pose*) di ciascuna camera. A partire da un insieme di immagini che inquadrano il soggetto da punti di vista diversi e parzialmente sovrapposti, un rilevatore di caratteristiche locali (tipicamente SIFT) individua e mette in corrispondenza punti salienti tra le immagini; da tali corrispondenze, attraverso la geometria epipolare e un processo di ottimizzazione globale noto come *bundle adjustment*, vengono stimati congiuntamente i parametri delle camere e le coordinate tridimensionali dei punti, ottenendo una nuvola sparsa. Nella seconda fase, la *Multi-View Stereo* (MVS) densifica la ricostruzione stimando mappe di profondità attraverso il confronto fotometrico tra viste multiple lungo le rette epipolari, così da generare una nuvola di punti densa. Da essa si ricava infine una mesh poligonale (ad esempio mediante ricostruzione di superficie di Poisson) sulla quale viene proiettata la texture estratta dalle fotografie (Figura 2.4).

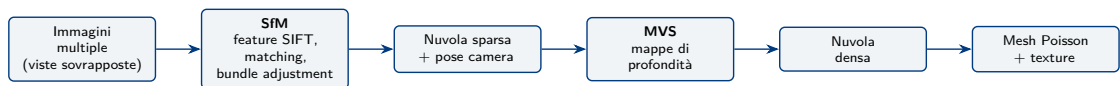


Figura 2.4: La pipeline fotogrammetrica classica: dalla stima delle pose (SfM) alla densificazione (MVS) fino alla mesh texturizzata.

Dal punto di vista matematico, la SfM è essenzialmente un problema di ottimizzazione non lineare. Le caratteristiche locali vengono individuate in uno spazio-scala (nel caso di SIFT, tramite differenze di gaussiane), così da risultare invarianti a scala e rotazione; le corrispondenze tra coppie di immagini sono quindi filtrate imponendo il vincolo epipolare, che lega geometricamente le proiezioni di uno stesso punto in due viste. Il *bundle adjustment* raffina infine congiuntamente i parametri

delle camere C_j e le coordinate dei punti tridimensionali X_i minimizzando l'errore di riproiezione complessivo:

$$\min_{\{C_j\}, \{X_i\}} \sum_{i,j} \|\mathbf{x}_{ij} - \pi(C_j, X_i)\|^2, \quad (2.1)$$

dove $\mathbf{x}_{ij}$ è la posizione osservata del punto i nell'immagine j e π la funzione di proiezione prospettica; il problema, fortemente non lineare, viene tipicamente risolto con metodi ai minimi quadrati smorzati come Levenberg–Marquardt. Nella fase di ricostruzione della superficie, l'algoritmo di Poisson interpreta la nuvola densa orientata come il campo del gradiente di una funzione indicatrice del solido, la cui stima si riconduce alla soluzione di un'equazione di Poisson: la mesh finale è l'isosuperficie estratta da tale funzione.

Pur raggiungendo un'elevata accuratezza metrica in condizioni ottimali, la fotogrammetria presenta limiti intrinseci legati al meccanismo di corrispondenza su cui si fonda. La stima della geometria presuppone infatti una sufficiente coerenza fotometrica tra immagini acquisite da punti di vista differenti, condizione che risulta generalmente valida per superfici opache e ben texturizzate, ma che viene meno in presenza di materiali riflettenti, trasparenti o privi di dettagli visivi, nei quali il *matching* tende a fallire producendo rumore o lacune nella ricostruzione. Inoltre, nelle pipeline fotogrammetriche tradizionali, l'informazione di illuminazione risulta spesso incorporata (*baked*) nella texture della geometria: il modello ottenuto non separa quindi in modo esplicito albedo, ombreggiatura e riflessioni, limitando la capacità di riprodurre fedelmente effetti luminosi dipendenti dal punto di vista e restituendo, in ambienti interattivi, una resa visiva spesso statica o poco naturale.

Un avanzamento significativo è stato introdotto nel 2020 con i *Neural Radiance Fields* (NeRF) proposti da Ben Mildenhall e colleghi [17]. Un NeRF rappresenta la scena come una funzione plenottica continua a cinque dimensioni, appresa da una rete neurale *Multi-Layer Perceptron*: a partire da una posizione spaziale (x, y, z) e da una direzione di vista (θ, φ), la rete stima la densità del volume e la radianza emessa, consentendo la sintesi di nuove viste tramite *volume rendering* differenziabile (Figura 2.5). Questo approccio ha permesso di raggiungere livelli di fotorealismo significativamente superiori ai metodi precedenti, modellando in modo convincente anche effetti dipendenti dal punto di vista, come variazioni speculari e riflessi moderati.

Più precisamente, il colore di ciascun pixel si ottiene integrando i contributi di radianza lungo il raggio $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ che attraversa la scena:

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t) \sigma(\mathbf{r}(t)) \mathbf{c}(\mathbf{r}(t), \mathbf{d}) dt, \quad T(t) = \exp\left(-\int_{t_n}^t \sigma(\mathbf{r}(s)) ds\right), \quad (2.2)$$

dove σ è la densità volumetrica stimata dalla rete, $\mathbf{c}$ la radianza dipendente dalla direzione e $T(t)$ la trasmittanza accumulata, ovvero la probabilità che il raggio giunga alla profondità t senza essere assorbito. In pratica l'integrale è approssimato per quadratura numerica, campionando un numero finito di punti lungo il raggio con una strategia gerarchica: una rete “coarse” individua le regioni a maggiore densità, sulle quali una rete “fine” concentra campioni aggiuntivi. Un ulteriore elemento chiave è il *positional encoding*: poiché le reti neurali tendono ad apprendere preferenzialmente funzioni a bassa frequenza, le coordinate di ingresso vengono proiettate in uno spazio di dimensione superiore mediante funzioni sinusoidali a frequenza crescente, consentendo al modello di rappresentare dettagli geometrici e tessiture ad alta frequenza.

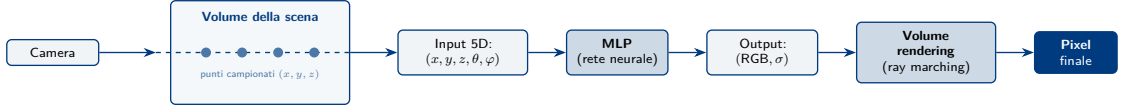


Figura 2.5: Il NeRF come funzione plenottica appresa: una rete MLP mappa posizione e direzione di vista in colore e densità; la sintesi di una nuova vista avviene per *volume rendering* lungo i raggi (*ray marching*).

Nella formulazione originaria, tuttavia, il costo computazionale del *ray marching* volumetrico comportava tempi di addestramento elevati e velocità di rendering incompatibili con applicazioni interattive in tempo reale su hardware consumer, limitandone l’impiego in contesti quali WebXR e tour immersivi.

2.2.2 Principi fondamentali e vantaggi prestazionali del 3D Gaussian Splatting (3DGS)

Nel 2023 Bernhard Kerbl e colleghi hanno introdotto il *3D Gaussian Splatting* (3DGS), un approccio che riduce significativamente il tradizionale compromesso tra qualità visiva e prestazioni computazionali [11]. Il 3DGS abbandona la rappresentazione implicita dei NeRF (in cui l’intera scena è codificata nei parametri di una

rete neurale) a favore di una rappresentazione esplicita, in cui la scena è descritta da un insieme di gaussiane tridimensionali anisotrope. Ciascuna gaussiana è definita da posizione, matrice di covarianza (che ne determina scala e orientamento), opacità e radianza dipendente dalla direzione di osservazione, modellata tramite armoniche sferiche. L'insieme delle gaussiane viene inizializzato a partire dalla nuvola sparsa prodotta dalla *Structure-from-Motion* e successivamente ottimizzato minimizzando la differenza tra le viste renderizzate e le immagini di riferimento. Durante l'addestramento, un meccanismo di controllo adattivo della densità clona, suddivide o rimuove gaussiane, addensandole nelle regioni in cui il dettaglio risulta insufficiente ed eliminando quelle ridondanti.

In termini formali, ciascuna primitiva è una gaussiana tridimensionale non normalizzata

$$G(\mathbf{x}) = \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right), \quad (2.3)$$

con centro $\boldsymbol{\mu}$ e matrice di covarianza Σ . Per garantire che Σ rimanga semidefinita positiva durante l'ottimizzazione per discesa del gradiente, essa viene parametrizzata come $\Sigma = R S S^\top R^\top$, dove S è una matrice diagonale di scala e R una rotazione espressa tramite quaternioni: la gaussiana è così un ellissoide liberamente orientabile e anisotropo. In fase di rendering, la covarianza viene trasportata nello spazio immagine mediante $\Sigma' = J W \Sigma W^\top J^\top$, con W trasformazione di vista e J jacobiano dell'approssimazione affine della proiezione prospettica; il colore di ciascun pixel risulta infine dalla composizione ordinata in profondità delle N gaussiane che vi si proiettano:

$$C = \sum_{i \in N} \mathbf{c}_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j), \quad (2.4)$$

dove α_i è l'opacità della gaussiana modulata dalla sua impronta ellittica e $\mathbf{c}_i$ il colore dipendente dalla vista, ricostruito dai coefficienti di armoniche sferiche associati alla primitiva.

A differenza dei NeRF, la visualizzazione non richiede *ray marching* volumetrico. Ogni gaussiana viene proiettata (*splatting*) sul piano immagine come un'impronta ellittica, e i contributi delle gaussiane sovrapposte sono composti mediante *alpha blending* ordinato in profondità attraverso un rasterizzatore differenziabile *tile-based*, che consente una parallelizzazione efficiente su GPU. Questo approccio permette

tempi di addestramento anche nell'ordine di minuti e rendering in tempo reale su GPU consumer di fascia medio-alta, senza richiedere l'inferenza di una rete neurale durante la visualizzazione (Figura 2.6).

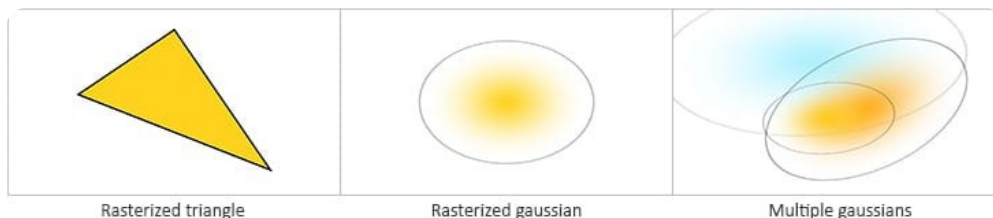


Figura 2.6: Dal triangolo rasterizzato allo *splatting*: una singola gaussiana proiettata come impronta ellittica e la composizione di più gaussiane sovrapposte tramite *alpha blending*.

La natura esplicita della rappresentazione costituisce un vantaggio rilevante per le pipeline impiegate nel presente lavoro: le nuvole di gaussiane possono essere esportate in formati standard, ripulite dagli artefatti, decimate, animate e integrate in workflow di produzione 3D e contenitori web. Le analisi comparative condotte sul patrimonio architettonico mostrano come il 3DGS offra un compromesso particolarmente efficace tra fedeltà visiva e interattività [18]. Va tuttavia sottolineato un limite importante: il 3DGS non garantisce l'accuratezza metrica richiesta dal rilievo conservativo, con errori geometrici che possono raggiungere l'ordine dei centimetri, e va pertanto inteso principalmente come strumento di restituzione visiva ed esperienziale piuttosto che di misurazione. È in questa prospettiva che viene impiegato nel presente lavoro: non per documentare metricamente i palcoscenici, ma per ricostruirne un'evocazione immersiva a nuvola di punti.

Il confronto sintetico tra le tre famiglie di tecniche è riassunto nella Tabella 2.1.

2.3 L'Intelligenza Artificiale nella ricostruzione tridimensionale

Le tecniche di ricostruzione analizzate nel §2.2 richiedono, per loro natura, un ampio insieme di immagini acquisite da punti di vista differenti e un significativo sforzo computazionale in fase di elaborazione. Più recentemente, si è affermata una linea di ricerca alternativa che mira a inferire la struttura tridimensionale della scena a

Tabella 2.1: Confronto sintetico tra le tre famiglie di tecniche di ricostruzione tridimensionale.

Criterio	Fotogrammetria	NeRF	3D Gaussian Splatting
Rappresentazione	Esplicita (mesh)	Implicita (rete MLP)	Esplicita (gaussiane 3D)
Output	Mesh + texture	Campo di radianza	Nuvola di gaussiane
Effetti view-dependent	Limitati (baked)	Sì (speculari/riflessi)	Sì (armoniche sferiche)
Tempo di training	Ore	Ore (formul. originaria)	Minuti
Rendering	Real-time	Lento (ray marching)	Real-time (rasterizzazione)
Accuratezza metrica	Alta	Media	Bassa (errori $\sim$ cm)
Editabilità/export	Alta (mesh standard)	Bassa	Alta (PLY, ripulibile)

partire da una singola immagine, sfruttando modelli di apprendimento profondo per la stima monoculare della geometria. È su questo paradigma che si basa la pipeline adottata nel presente lavoro, applicata inoltre a immagini equirettangolari, un dominio che introduce specifiche criticità legate alla distorsione della proiezione.

2.3.1 Algoritmi AI per la stima monoculare della profondità (Focus: Depth Anything V2)

Il principale ostacolo nel ricostruire una scena tridimensionale a partire da una singola immagine è la stima della profondità di ciascun pixel, problema noto come *Monocular Depth Estimation* (MDE). La difficoltà deriva dalla natura stessa della proiezione prospettica, che comprime la scena tridimensionale su un piano bidimensionale eliminando l'informazione sulla distanza. Di conseguenza, una stessa immagine può essere compatibile con infinite configurazioni geometriche differenti: un oggetto piccolo e vicino e uno grande e lontano, ad esempio, possono produrre esattamente la stessa proiezione sul piano immagine.

Non essendo possibile risolvere il problema attraverso la sola geometria, la stima

della profondità richiede necessariamente l'introduzione di conoscenze a priori. In modo analogo al sistema visivo umano, che inferisce la profondità sfruttando indizi monoculari appresi con l'esperienza (come dimensione relativa degli oggetti, occlusioni, gradienti di texture e prospettiva), i più recenti *foundation model* apprendono tali regolarità statistiche da grandi quantità di dati, imparando ad associare a ciascuna immagine la configurazione tridimensionale più plausibile.

Depth Anything V2 [9] adotta in particolare un'architettura basata su un encoder DINOv2 e un decoder DPT. Il modello viene addestrato combinando immagini sintetiche annotate con grandi quantità di immagini reali pseudo-etichettate, sfruttando un processo di distillazione della conoscenza da un modello *teacher* di maggiore capacità. Questo approccio consente di ottenere mappe di profondità robuste e ricche di dettaglio a partire da una singola immagine RGB.

Va tuttavia precisato che l'output di Depth Anything V2 corrisponde a una profondità relativa (*affine-invariant*): il modello cattura correttamente le relazioni di profondità tra le diverse regioni della scena, ma non ne restituisce la scala metrica assoluta. Per convertire la mappa di profondità in coordinate tridimensionali correttamente proporzionate è quindi necessario introdurre una normalizzazione oppure un riferimento di scala esterno.

2.3.2 Adattamento per proiezioni equirettangolari: estensione di modelli prospettici a formati immersivi (Apple SHARP e utilizzo delle Cube-map)

Per la generazione di rappresentazioni gaussiane tridimensionali a partire da una singola immagine, il modello di riferimento considerato in questo lavoro è SHARP [10]. A partire da una singola immagine RGB, SHARP regredisce in un'unica passata *feedforward* i parametri di una rappresentazione basata su *Gaussian Splats* in scala metrica, renderizzabile in tempo reale. Più che ricostruire una geometria esplicita e completa della scena nel senso tradizionale del termine, il modello apprende una rappresentazione tridimensionale plausibile ottimizzata per la sintesi di viste vicine (*novel view synthesis*), consentendo di generare prospettive leggermente traslate rispetto all'immagine di input con elevata coerenza visiva.

Come molti modelli monoculari, tuttavia, SHARP è addestrato su fotocamere prospettiche (*pinhole*) e non supporta nativamente immagini equirettangolari o fisheye. UniSHARP [19] affronta questo limite generalizzando l'approccio a sistemi ottici arbitrari. L'idea centrale consiste nel disaccoppiare la ricostruzione dalla griglia bidimensionale dei pixel (intrinsecamente dipendente dal modello di camera) e nell'esprimere invece la posizione di ciascuna primitiva gaussiana rispetto a un fascio di raggi definito in coordinate sferiche, secondo una rappresentazione detta *ray-distance*. In questo modo, anche immagini panoramiche a 360° possono essere convertite direttamente in rappresentazioni gaussiane continue senza necessità di scomposizione in viste prospettiche separate.

SHARP produce tuttavia rappresentazioni affidabili soprattutto per variazioni limitate del punto di vista attorno alla vista di input: traslazioni ampie o regioni inizialmente occluse possono generare artefatti o geometrie allucinate. Inoltre, UniSHARP rappresenta ancora un risultato di ricerca recente e non è attualmente integrato in strumenti di produzione consolidati. Per questa ragione, la pipeline adottata nel presente lavoro impiega una soluzione più pragmatica basata sulla repository open-source SHARP_360_to_Splat [20]. Il panorama equirettangolare viene dapprima suddiviso in più viste prospettiche, analogamente a una *cubemap* (Figura 2.7); SHARP genera quindi le gaussiane per ciascuna vista in modo indipendente, e le nuvole risultanti vengono successivamente allineate e fuse in una scena unificata. L'elaborazione separata delle diverse facce può tuttavia introdurre incoerenze di scala e discontinuità ai bordi (*seam artifacts*); per attenuare tali problemi, la pipeline sfrutta DA360 [21] come riferimento di scala globale e include una fase di riallineamento delle regioni sovrapposte. Il dettaglio operativo di questo flusso e la sua integrazione nella pipeline complessiva saranno discussi nel §3.3.



Figura 2.7: Dalla proiezione equirettangolare alla *cube-map*: il panorama a 360° dell'Auditorium del Parco (a sinistra) e la sua scomposizione in facce prospettiche (a destra), primo passaggio della pipeline SHARP_360_to_Splat.

Capitolo 3

Metodologia e Sviluppo della Pipeline Ibrida

Questo capitolo descrive nel dettaglio la pipeline ibrida sviluppata per la produzione dei documentari immersivi, dall'acquisizione sul set fino all'integrazione nell'interfaccia di fruizione. La Figura 3.1 ne offre una visione d'insieme: il flusso principale segue la lavorazione dei tre sistemi di ripresa fino ai master stereoscopici a 360°, mentre il ramo parallelo illustra la ricostruzione degli ambienti teatrali a nuvola di punti destinati al menu immersivo; i due flussi convergono nell'ambiente di authoring 3DVista. Le sezioni seguenti approfondiscono ciascuna fase dello schema.

3.1 Acquisizione dei dati e setup di ripresa sul set

All'avvio delle attività, il progetto itARTS era già in corso da tempo: si è resa quindi necessaria una fase preliminare di studio e analisi dell'iniziativa nel suo complesso e delle singole opere, così da comprenderne finalità, linguaggi e tempi di sviluppo. A questa attività di orientamento si è affiancata una parallela fase di pianificazione delle riprese e di presa di contatti con gli enti e i protagonisti coinvolti, condotta dal prof. Sburlati, che ha permesso di definire calendari, accessi e modalità operative sul set.

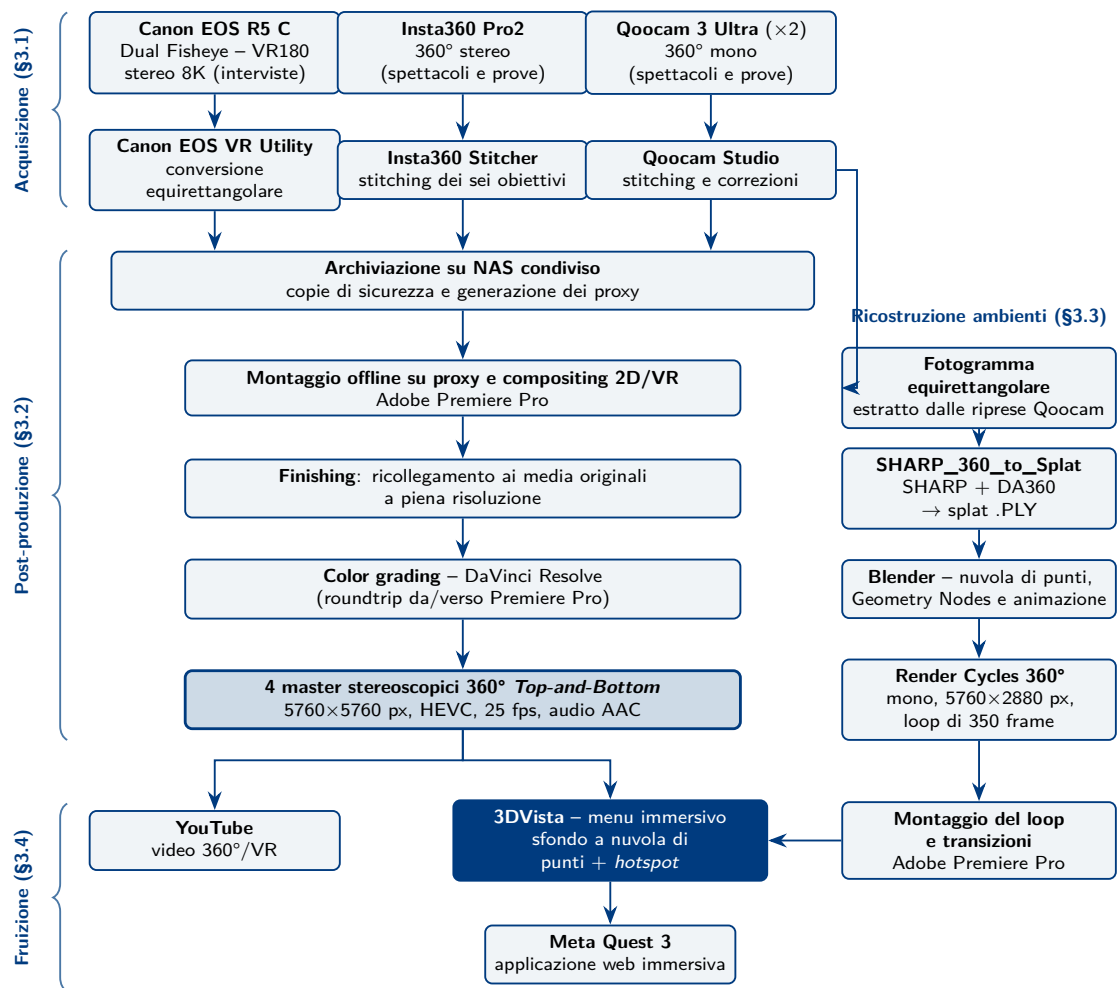


Figura 3.1: Schema d'insieme della pipeline ibrida: il flusso di produzione dei documentari (a sinistra) e il ramo di ricostruzione degli ambienti a nuvola di punti (a destra) convergono nel menu immersivo realizzato in 3DVista. Per chiarezza sono omessi i dispositivi di riserva (B camera e audio di sicurezza).

L'attività di ripresa si è articolata in quindici giornate distribuite nell'arco di circa tre mesi. La maggior parte delle riprese si è svolta all'Aquila, tra l'Accademia di Belle Arti e il Conservatorio "A. Casella", mentre quattro giornate hanno avuto luogo ad Annecy, al seguito della tournée internazionale del progetto. Il materiale raccolto è considerevole: ventisette interviste, oltre cinquanta ore di girato e un volume complessivo di circa ventisette terabyte di dati (Figura 3.2).

L'impianto di acquisizione rispecchiava la natura ibrida della pipeline, assegnando a ciascuna tipologia di contenuto lo strumento più adatto. Le interviste ai protagonisti sono state interamente affidate alla Canon R5 C equipaggiata con



Figura 3.2: La troupe al lavoro durante le riprese delle interviste ai protagonisti del progetto.

ottica Dual Fisheye (§3.1.1.1), affiancata da una DJI Osmo Pocket impiegata come B camera. Per la documentazione degli ambienti, degli spettacoli e delle prove si è invece adottato un approccio multi-camera, impiegando simultaneamente tutti i sistemi a disposizione: trattandosi di eventi in larga parte irripetibili, come le rappresentazioni dal vivo e le prove, si è scelto di moltiplicare i punti di vista per non perdere alcun momento significativo. La Tabella 3.1 riassume i sistemi impiegati e i relativi flussi di pre-elaborazione.

Questa strategia ha però dovuto misurarsi con un vincolo specifico legato alle rappresentazioni aperte al pubblico, durante le quali non era possibile collocare camere sul palco né in posizioni troppo esposte alla vista degli spettatori. Le riprese ravvicinate sono state perciò concentrate sulle prove generali, mentre nel corso degli spettacoli le camere sono state disposte in punti più defilati, così da non interferire con la fruizione della scena.

3.1.1 Riprese stereoscopiche

Riprese a 180° (Canon R5 C e ottica Dual Fisheye)

Le interviste ai protagonisti, che costituiscono il fulcro narrativo della documentazione, sono state realizzate con una Canon EOS R5 C equipaggiata con l'ottica Canon RF5.2mm F2.8L Dual Fisheye, il cui principio di funzionamento è stato

Tabella 3.1: Quadro riassuntivo dei sistemi di acquisizione impiegati sul set e dei relativi flussi di pre-elaborazione.

Sistema	Formato	Impiego principale	Pre-elaborazione
Canon EOS R5 C con ottica Dual Fisheye	VR180 stereoscopico, 8K, 25 fps, Canon Log 3	Interviste ai protagonisti	Canon EOS VR Utility
Insta360 Pro2	360° stereoscopico	Spettacoli e prove nell'ap- proccio multi-camera	Insta360 Stitcher
Qoocam 3 Ultra (×2)	360° monoscopico	Punti di vista ravvicinati; fotogrammi sorgente per la ricostruzione degli ambienti (§3.3)	Qoocam Studio
DJI Osmo Pocket	2D	B camera di riserva per le interviste	—
Radiomicrofoni lavalier Rode; Zoom H3-VR	Audio	Voce delle interviste; trac- cia ambientale di sicurezza	—

descritto nel §2.1.2. L'impiego di questo sistema risponde a una precisa esigenza espressiva: la percezione stereoscopica della profondità, unita alla collocazione frontale dello spettatore rispetto al soggetto, consente di valorizzare la presenza fisica dell'intervistato e di rafforzare il coinvolgimento percettivo. A ciò si aggiunge un vantaggio operativo rilevante: l'ampia copertura dell'ottica consente all'intervistatore di posizionarsi immediatamente dietro la camera, preservando una direzione dello sguardo naturale e favorendo l'impressione, per chi fruisce del contenuto, di essere direttamente coinvolto nell'interazione.

Le riprese sono state effettuate in risoluzione 8K con profilo colore Canon Log 3, così da preservare la massima latitudine di posa per la successiva fase di correzione colore. L'esposizione è stata gestita in manuale: con un frame rate di 25 fotogrammi al secondo, il tempo di otturazione è stato impostato secondo la regola dei 180°, corrispondente a circa 1/50 di secondo, così da ottenere un *motion blur* naturale, mentre la sensibilità ISO è stata adattata di volta in volta alle condizioni di luce dei diversi ambienti; le interviste, in particolare, si sono avvalse della sola luce ambientale degli spazi, senza un impianto di illuminazione dedicato.

Sul piano dell'attrezzatura, la camera è stata stabilizzata su un treppiede Cartoni dotato di testa fluida. Per escludere il cavalletto dall'inquadratura, problema tipico delle riprese immersive, in cui il campo ripreso abbraccia un intero emisfero, è stato impiegato un braccio di estensione; il conseguente spostamento del baricentro è stato compensato con un contrappeso (*sandbag*), a garanzia della stabilità del sistema. Il controllo della messa a fuoco e dell'inquadratura (reso più delicato dall'ampiezza del campo del Dual Fisheye) è stato affidato a un monitor esterno. A integrazione della R5 C è stata impiegata una DJI Osmo Pocket come B Camera, a tutela del materiale in caso di imprevisti.

L'audio delle interviste è stato acquisito tramite radiomicrofoni lavalier wireless Rode Mic Pro, affiancati da un registratore ambisonico Zoom H3-VR, posizionato in prossimità della camera, che ha fornito una traccia ambientale a quattro canali impiegata come pista secondaria e di sicurezza.



Figura 3.3: L'attrezzatura impiegata per le interviste: la Canon EOS R5 C (a sinistra), la DJI Osmo Pocket usata come B camera (al centro) e il registratore ambisonico Zoom H3-VR (a destra).

La realizzazione delle interviste ha comportato alcune criticità, di natura tecnica e produttiva. Sul piano tecnico, l'ottica Dual Fisheye ha richiesto particolare attenzione all'allineamento e alla messa a fuoco delle due lenti: eventuali discrepanze tra di esse si traducono infatti in difetti stereoscopici, e il controllo costante dell'immagine su un monitor esterno si è rivelato indispensabile per individuarli e correggerli tempestivamente. La ripresa in 8K ha inoltre imposto una gestione oculata dell'alimentazione e dello storage, dato l'elevato consumo energetico e l'ingente volume di dati generato. Sul piano produttivo, il vincolo principale è stato il tempo: le interviste dovevano incastrarsi tra le prove e gli impegni dei protagonisti, e un contesto fatto in larga parte di eventi irripetibili lasciava un

marginare ridotto, imponendo una pianificazione flessibile e la prontezza a riprendere non appena le condizioni lo consentivano.

Riprese a 360° (Insta360 Pro2)

La Insta360 Pro2 è stata impiegata, insieme alle altre camere, per la ripresa degli spettacoli, nell'ambito dell'approccio multi-camera descritto in apertura. Le sue dimensioni, tuttavia, ne hanno condizionato il posizionamento: a differenza delle Qoocam (estremamente compatte e perciò facilmente occultabili in vari punti dello spazio senza interferire con la scena o con il pubblico), la Pro2 risultava troppo ingombrante per essere collocata in prossimità del palco o dei musicisti durante le rappresentazioni pubbliche. Le riprese ravvicinate con

questa camera sono state perciò realizzate esclusivamente in occasione delle prove generali, quando la presenza di un corpo macchina visibile non comprometteva la fruizione dello spettacolo; durante le rappresentazioni vere e proprie, invece, la Pro2 è stata collocata in posizioni più defilate.

Sul piano tecnico, la Insta360 Pro2 è stata impiegata in modalità stereoscopica, per restituire la profondità della scena, e mantenuta sempre fissa su uno stativo, senza ricorrere né alla stabilizzazione né al sistema di monitoraggio remoto FarSight. L'audio è stato in genere acquisito direttamente dai microfoni integrati della camera, senza una registrazione separata. La cucitura dei flussi provenienti dai sei obiettivi è stata infine effettuata in post-produzione con il software Insta360 Stitcher.

Diversamente da quanto spesso riportato per le riprese sferiche, lo stitching non ha costituito una criticità rilevante: poiché di norma nessun soggetto si trovava in prossimità della camera, le cuciture tra gli obiettivi non hanno generato artefatti significativi. Si è resa comunque necessaria qualche piccola correzione di orizzonte e di verticalità sui flussi provenienti sia dalla Insta360 Pro2 sia dalle Qoocam 3 Ultra, effettuata direttamente in fase di stitching all'interno dei rispettivi software (Insta360 Stitcher e Qoocam Studio), così da compensare i lievi disallineamenti dovuti al posizionamento delle camere sul set. La ripresa a 360°, in cui l'intero spazio



Figura 3.4: Insta360 Pro2.

circostante rientra nel campo, impedisce tuttavia di allestire un set tradizionale o di mantenere operatori in scena; per ovviare a questo limite comune anche alle Qoocam le camere venivano avviate da remoto tramite l'apposita applicazione, così che gli operatori potessero allontanarsi dall'inquadratura. La principale criticità è stata piuttosto la gestione dei file: la mole di dati generata dalle riprese a 360° ad alta risoluzione ha richiesto una specifica strategia di archiviazione e backup, affrontata nel §3.2.

3.1.2 Riprese monoscopiche a 360° (Qoocam 3 Ultra)

Le riprese monoscopiche a 360° sono state realizzate con due Qoocam 3 Ultra, impiegate per valorizzare la dimensione immersiva delle opere. L'adozione di due unità ha permesso di catturare simultaneamente più punti di vista, mentre le dimensioni estremamente contenute di queste camere hanno consentito di collocarle in prossimità dei performer, in posizioni difficilmente accessibili ad apparecchiature più ingombranti. Si sono così ottenuti punti di vista insoliti e di forte impatto, come una soggettiva del direttore d'orchestra, ponendo la camera di fronte all'ensemble, o inquadrature da sotto il palco. Come per la Insta360 Pro2, l'avvio delle camere avveniva da remoto tramite l'applicazione dedicata e l'audio era acquisito direttamente dai dispositivi.



Figura 3.5: Kandao Qoocam 3 Ultra.

Le riprese effettuate con la Qoocam costituiscono inoltre la base di partenza per la ricostruzione tridimensionale degli ambienti descritta nel §3.3: è da questi fotogrammi a 360° che sono state estratte le immagini equirettangolari impiegate per generare gli splat.

La principale criticità ha riguardato il comportamento delle camere in condizioni di scarsa illuminazione: i sensori delle Qoocam soffrivano in modo marcato gli ambienti bui, restituendo immagini affette da rumore. Poiché la maggior parte degli spazi presentava condizioni di luce complesse tipiche del contesto teatrale, con forti contrasti tra zone illuminate e fondali in ombra, si è reso necessario un intervento di riduzione del rumore in fase di post-produzione.

3.2 Post-produzione e organizzazione dei flussi di lavoro

3.2.1 Gestione dei dati, stitching e ottimizzazione dei file ad alta risoluzione

L'ingente volume di materiale prodotto, dell'ordine dei ventisette terabyte, ha imposto una rigorosa strategia di gestione dei dati fin dal set. Al termine di ogni giornata di ripresa il girato veniva duplicato in circa tre copie su altrettanti hard disk distinti, così da tutelare il materiale da eventuali perdite o guasti.

Ciascun sistema di ripresa richiedeva inoltre una specifica fase di elaborazione preliminare, volta a convertire i flussi grezzi nel formato equirettangolare. Le riprese stereoscopiche a 180° della Canon R5 C sono state elaborate con Canon EOS VR Utility; i flussi provenienti dai sei obiettivi della Insta360 Pro2 sono stati cuciti con Insta360 Stitcher (§3.1.1.2); le riprese monoscopiche a 360° delle Qoocam 3 Ultra sono state processate con Qoocam Studio.

Al fine di ridurre il carico computazionale durante il montaggio, ciascuna clip è stata transcodificata in una versione proxy a risoluzione e bitrate ridotti, utilizzata nella fase di editing offline e descritta nel paragrafo seguente.

3.2.2 Workflow collaborativo e archiviazione condivisa tra diverse postazioni di calcolo

I file originali e quelli risultanti dallo stitching sono stati archiviati su un NAS veloce condiviso di alta capacità nello studio di Motion Pixel, utilizzato come repository centralizzata dell'intero materiale di produzione. Per ottimizzare il workflow di montaggio, l'editing è stato condotto prevalentemente su file proxy, copiati su un'unità di archiviazione locale, riducendo così la necessità di accedere continuamente ai media sorgente ad alta risoluzione e alleggerendo il carico di trasferimento dati. Solo nella fase di *finishing* le sequenze montate sono state ricollegate ai file originali, così da operare direttamente sul materiale a piena risoluzione.

La pipeline di color grading ha seguito un flusso separato. Una volta completato il montaggio in Adobe Premiere Pro, la timeline veniva esportata in formato non compresso e ad alta risoluzione, priva di elementi grafici sovrapposti, per essere elaborata in DaVinci Resolve. Il materiale corretto cromaticamente veniva infine reimportato in Premiere per l'assemblaggio finale e l'esportazione del master definitivo.

3.2.3 Integrazione dei contributi bidimensionali delle opere nelle scene a 180°

Come osservato nel §2.1.2, il formato a 180° lascia inutilizzata la porzione di sfera posta alle spalle dell'osservatore. Tale spazio è stato sfruttato per integrare, all'interno delle scene stereoscopiche, materiali visivi specifici di ciascun documentario, con l'obiettivo di arricchirne il contesto e rafforzarne la dimensione narrativa.

La natura dei contributi variava in funzione del contenuto. Per *Dei Maestri Dipintori* (spettacolo che mette in scena la teatralità di tre celebri dipinti di Paolo Uccello, Giandomenico Tiepolo e Alberto Burri) sono state utilizzate le riproduzioni delle opere pittoriche originali. Per *Donne all'Opera: Croce e Delizia* sono state invece integrate le animazioni generate mediante intelligenza artificiale e realizzate per il LED wall dello spettacolo (una per ciascuna aria) insieme a maquette, bozzetti scenografici e studi per i costumi. In *Shared*, infine, sono stati impiegati i segni grafici già utilizzati come asset visivi dell'installazione e dell'esperienza in Realtà Virtuale.

L'inserimento di questi materiali è avvenuto in fase di post-produzione attraverso operazioni di compositing. L'intera lavorazione (dal montaggio al compositing) è stata eseguita in Adobe Premiere Pro (Figura 3.7); la scelta di questo software è stata determinata dal suo solido supporto nativo agli strumenti VR, che ha consentito di gestire agevolmente i metadati equirettangolari e di visualizzare in tempo reale l'ambiente sferico durante l'editing. I diversi contributi sono stati collocati nelle aree della sfera non occupate dalla ripresa VR180 e disposti dietro il punto di osservazione come superfici visive sospese nello spazio immersivo. Mentre dipinti, bozzetti, maquette e segni grafici sono stati integrati come elementi statici, le animazioni relative alle arie di *Donne all'Opera*, già disponibili in formato video,

sono state mantenute in movimento, introducendo una componente dinamica che contribuisce a rendere più viva la scena immersiva (Figura 3.6).



Figura 3.6: Fotogramma monoscopico di *Donne all'Opera*: al centro la scena VR180 con l'intervista, mentre nell'emisfero posteriore (ai lati) sono integrati i contributi bidimensionali dell'opera.

Oltre ai tre documentari dedicati alle singole opere, è stato realizzato un quarto video di sintesi destinato a raccontare le tre opere insieme e, in generale, l'intero progetto itARTS. Questo contenuto è stato assemblato estraendo e ricombinando i passaggi più significativi delle tre produzioni principali, mantenendone intatto l'impianto visivo immersivo ma adottando un ritmo di montaggio più dinamico, funzionale a un racconto corale e riassuntivo.

3.2.4 Montaggio e grammatica delle transizioni immersive

La fase di montaggio dei documentari immersivi ha seguito una specifica grammatica di editing VR, strutturata per coniugare l'efficacia narrativa dell'esperienza con il comfort visivo dello spettatore.

La struttura narrativa adotta un approccio di **storytelling guidato dalle interviste** (*interview-driven storytelling*) basato su un **montaggio corale**. Senza l'utilizzo di una voce narrante fuori campo, il racconto è costruito unendo le testimonianze dirette di docenti, registi, studenti e artisti. Le risposte degli

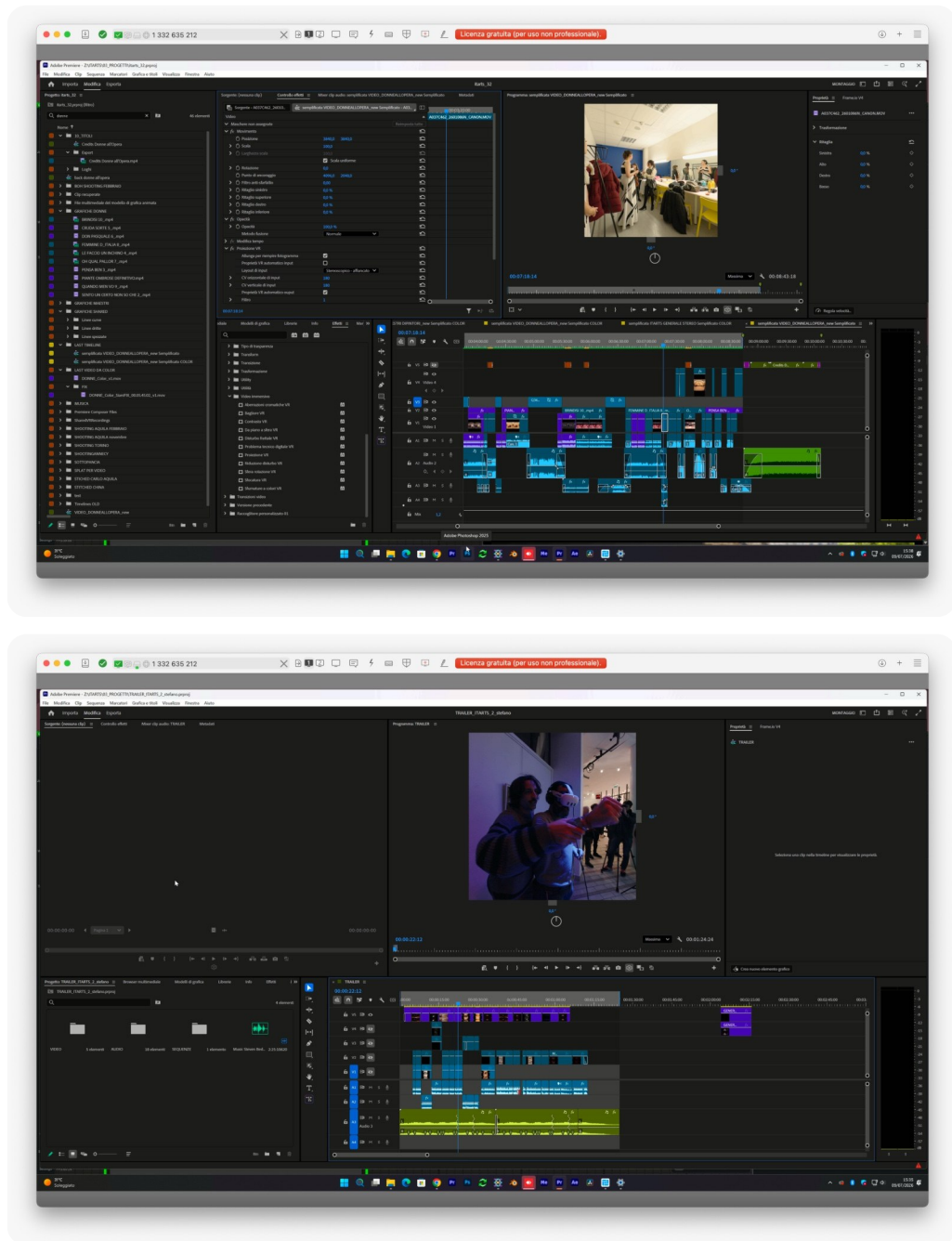


Figura 3.7: La post-produzione in Adobe Premiere Pro: in alto il *compositing* dei contributi 2D/VR nel documentario (con gli effetti di proiezione immersiva e le grafiche delle arie); in basso il montaggio del video di presentazione.

intervistati si alternano e si integrano tra loro, creando un filo conduttore che fa progredire la narrazione e restituisce i diversi punti di vista sul progetto.

Il ritmo di montaggio (*pacing*) è caratterizzato da inquadrature lunghe: la durata media delle clip a 360 gradi varia tra i 15 e i 20 secondi, mentre per le interviste in VR180 si sono mantenuti blocchi temporali ancora più estesi. Si è evitato l'inserimento di clip inferiori agli 8-10 secondi. Questa scelta consente allo spettatore di orientarsi all'interno dello spazio sferico e osservare l'ambiente senza interruzioni troppo frequenti, favorendo l'immedesimazione nella scena, coerentemente con la necessità, evidenziata dalla letteratura sulla regia immersiva, di ripensare il ritmo del montaggio tradizionale a favore dei tempi di esplorazione dello spettatore [33].

Per il passaggio tra le clip si è scelto di utilizzare prevalentemente lo stacco netto (*hard cut*), limitando l'uso delle dissolvenze incrociate solo ad alcuni raccordi specifici per evitare transizioni grafiche ripetitive. Si tratta di una scelta diversa da quella di altri autori del settore (Felix & Paul Studios privilegiano ad esempio dissolvenze lente, percepite come un aprirsi e chiudersi degli occhi [38]), ma motivata dalla natura documentaristica del materiale: lo stacco netto preserva la neutralità del racconto ed evita che una transizione ricorrente si trasformi in un manierismo visivo. Anche il passaggio tra il formato VR180 delle interviste e il 360 gradi degli spettacoli avviene tramite stacco netto: in questo modo, lo spazio alle spalle dello spettatore passa istantaneamente dall'essere occupato dai contributi bidimensionali all'apertura completa sullo spazio teatrale a 360 gradi. Lo stacco netto assorbe inoltre il passaggio tra sorgenti stereoscopiche e monoscopiche: l'attenuazione della profondità binoculare coincide così con un cambio complessivo di scena, risultando percettivamente assai meno saliente (le implicazioni di questa coesistenza sono discusse nel §4.2).

Allo stacco tra le inquadrature, si è gestito l'orientamento orizzontale (*yaw*) allineando in post-produzione (tramite l'effetto *VR Projection* di Adobe Premiere Pro) l'asse principale della nuova clip con la direzione dello sguardo dello spettatore. In questo modo, il soggetto principale o l'azione centrale della nuova scena si trova nel cono visivo frontale (entro circa $\pm 30^\circ$ rispetto al centro), evitando che l'utente debba ruotare bruscamente il collo per individuare il punto di interesse; la pratica

risponde direttamente alla tendenza degli spettatori a mantenere un orientamento prevalentemente frontale, documentata dagli studi sulla guida dell'attenzione richiamati nel §2.1.3 [16, 34].

Infine, le inquadrature sono state mantenute rigorosamente statiche su stativo, escludendo movimenti di macchina virtuali o zoom digitali. Eventuali lievi inclinazioni del treppiede sul set sono state corrette in fase di *stitching* o tramite filtri in post-produzione, garantendo che la linea dell'orizzonte rimanesse perfettamente livellata rispetto al visore per prevenire disturbi da chinetosi.

3.3 Ricostruzione volumetrica e spaziale degli scenari performativi

La ricostruzione tridimensionale degli ambienti, destinata a costituire gli sfondi a nuvola di punti del menu di fruizione, è stata realizzata a partire da fotografie e da fotogrammi estratti dalle riprese effettuate con la Qoocam 3 Ultra. Il tratto distintivo dell'approccio adottato è l'impiego di una **singola immagine equirettangolare** per ciascuna ricostruzione: a differenza delle tecniche tradizionali, che richiedono numerosi scatti da punti di vista differenti, ogni scenario è stato ricostruito a partire da un solo fotogramma a 360°.

Sono stati ricostruiti tre palcoscenici, tutti situati all'Aquila: il palco dell'Accademia di Belle Arti, quello della *L'Aquila Temporary Concert Hall* (la sala concerti temporanea in tubi di cartone progettata da Shigeru Ban) e quello dell'Auditorium del Parco, l'auditorium in legno firmato da Renzo Piano; le ultime due sedi sono legate all'attività del Conservatorio "A. Casella".

Sul piano computazionale, la pipeline è stata eseguita su una postazione equipaggiata con una GPU NVIDIA RTX 4090 e 128 GB di memoria RAM; va tuttavia sottolineato come il workflow risulti poco esigente e possa essere eseguito agevolmente anche su hardware di fascia inferiore. La generazione di un singolo splat grezzo richiedeva circa trenta secondi, un tempo che, confrontato con le ore necessarie alle pipeline fotogrammetriche o basate su NeRF (§2.2), evidenzia il principale vantaggio pratico dell'approccio adottato.

3.3.1 Sperimentazione e confronto delle metodologie di reality capture

Fin dall'inizio l'obiettivo è stato la ricostruzione degli ambienti tramite *3D Gaussian Splatting*; sono state pertanto escluse a priori sia la fotogrammetria tradizionale, orientata alla generazione di mesh, sia i Neural Radiance Fields, concentrando la sperimentazione sui diversi modi di ottenere uno splat gaussiano. Prima di approdare alla soluzione definitiva sono stati valutati due approcci basati su immagini multiple.

Il primo ricalca la pipeline classica di generazione di uno splat a partire da più fotografie. A partire da una serie di scatti a 360° dei palcoscenici, il software AliceVision [22] è stato impiegato per estrarre dalle immagini equirettangolari un insieme di viste prospettiche; queste venivano poi elaborate con COLMAP [23] per il calcolo delle pose della camera (*Structure-from-Motion*) e infine date in input a Brush [24], un addestratore di Gaussian Splatting open-source, per la generazione vera e propria della nuvola di gaussiane.

Il secondo approccio muoveva dall'idea di sfruttare direttamente il visore per acquisire gli ambienti. In quel periodo Meta aveva presentato Hyperscape [25], una tecnologia proprietaria capace di trasformare una stanza reale in un ambiente fotorealistico navigabile in Realtà Virtuale sotto forma di Gaussian Splatting. Sulla stessa scia è stata sperimentata SplatareaScan [26], un'applicazione per il visore Meta Quest che consente di acquisire direttamente gli ambienti e di derivarne gli splat gaussiani, secondo un principio simile alla combinazione COLMAP–Brush ma in forma più automatizzata.

Entrambi gli approcci, tuttavia, condividevano i limiti propri delle tecniche basate su immagini multiple: la necessità di numerosi scatti, un'elevata richiesta di potenza di calcolo e una qualità dei risultati non sempre garantita; il secondo, in particolare, manifestava difficoltà evidenti negli ambienti poco illuminati. Queste considerazioni hanno motivato il passaggio a un approccio radicalmente più snello, basato sulla generazione dello splat da una singola immagine, descritto nella sezione seguente.

3.3.2 Stima della profondità (Depth Anything V2 e Apple SHARP) e generazione degli splat gaussiani / nuvola di punti

Ogni ambiente è stato ricostruito a partire da un singolo fotogramma equirettangolare estratto dalle riprese a 360° della Qoocam. L'immagine veniva fornita in input alla pipeline SHARP_360_to_Splat (Figura 3.8), che ne suddivide il contenuto in viste prospettiche prima di applicare il modello SHARP, incaricato di generare gli splat gaussiani per ciascuna faccia. Nella configurazione adottata il panorama è stato scomposto in **quattro facce**: tale valore si è rivelato il miglior compromesso, poiché aumentare il numero di viste incrementava la sovrapposizione (*overlap*) tra di esse e ne degradava la qualità complessiva. La normalizzazione della scala tra le facce è stata affidata a DA360, integrato nella repository, mentre le opzioni di upscaling (SeedVR2) e di *deblur* sono state mantenute disattivate. L'output era costituito da nuvole di splat gaussiani in formato .PLY, generate (come già osservato) nell'ordine dei trenta secondi ciascuna.

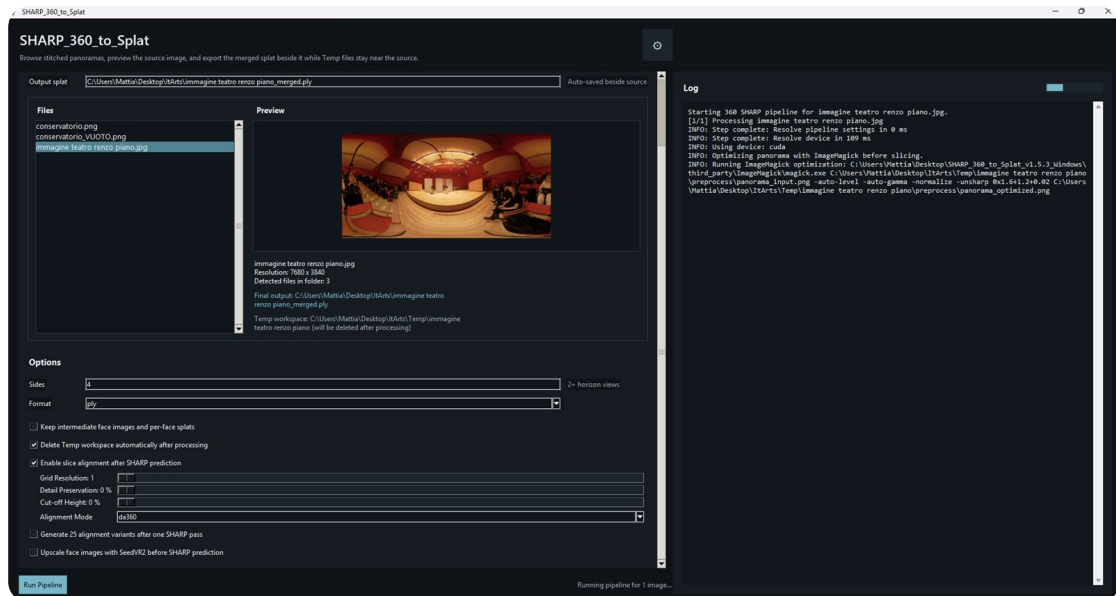


Figura 3.8: L'interfaccia della pipeline *SHARP_360_to_Splat*: il panorama equirettangolare in input, la scomposizione in quattro facce (*Sides*) e la normalizzazione della scala con DA360 (*Alignment Mode*), con esportazione dello splat in formato .PLY.

La rappresentazione ottenuta presentava due caratteristiche significative. In

primo luogo, la densità della nuvola di punti non era uniforme: risultava molto elevata in prossimità del punto di ripresa e decresceva progressivamente con la distanza. In secondo luogo, in corrispondenza dei poli (le zone immediatamente al di sotto e al di sopra della camera) si generava un'area scoperta di forma quadrata, conseguenza diretta della suddivisione del panorama in quattro viste. Entrambe le caratteristiche sono state affrontate nella successiva fase di ottimizzazione in Blender (§3.3.3).

3.3.3 Ottimizzazione e modellazione degli asset tridimensionali in Blender

Le nuvole di splat prodotte dalla pipeline sono state importate in Blender tramite l'importatore nativo per il formato Stanford PLY. In questo modo lo splat non veniva interpretato come un insieme di gaussiane renderizzabili, bensì come una **nuvola di punti** costituita dai vertici del file, successivamente elaborata e resa visivamente attraverso gli strumenti procedurali del software.

La resa e l'animazione della nuvola sono state costruite combinando un albero di *Geometry Nodes* con una serie di modificatori. All'interno dei *Geometry Nodes* la geometria in ingresso veniva convertita in punti mediante il nodo *Mesh to Points* (con un raggio di 0,008 m); la posizione di ciascun punto era quindi perturbata da un nodo *Set Position*, il cui *offset* era pilotato da una *Noise Texture* quadridimensionale, la cui quarta coordinata (W), variando nel tempo, produceva un movimento fluido e continuo dei punti, per poi assegnare all'insieme un **materiale emissivo**, così da farlo risaltare come elemento luminoso nello spazio. Il materiale era un semplice *Principled BSDF* con una componente di emissione di colore azzurro e intensità pari a circa 3,1, che restituiva alla nuvola l'aspetto luminoso e uniforme desiderato (Figura 3.9).

A questa animazione di base si sono aggiunti due modificatori che ne definivano il carattere complessivo: un modificatore *Simple Deform* in modalità *Twist*, responsabile di una lenta rotazione della nuvola, e un modificatore *Lattice*, la cui gabbia di deformazione era a sua volta animata da un modificatore *Wave*, in modo da imprimere all'insieme un moto ondulatorio. La combinazione di questi elementi

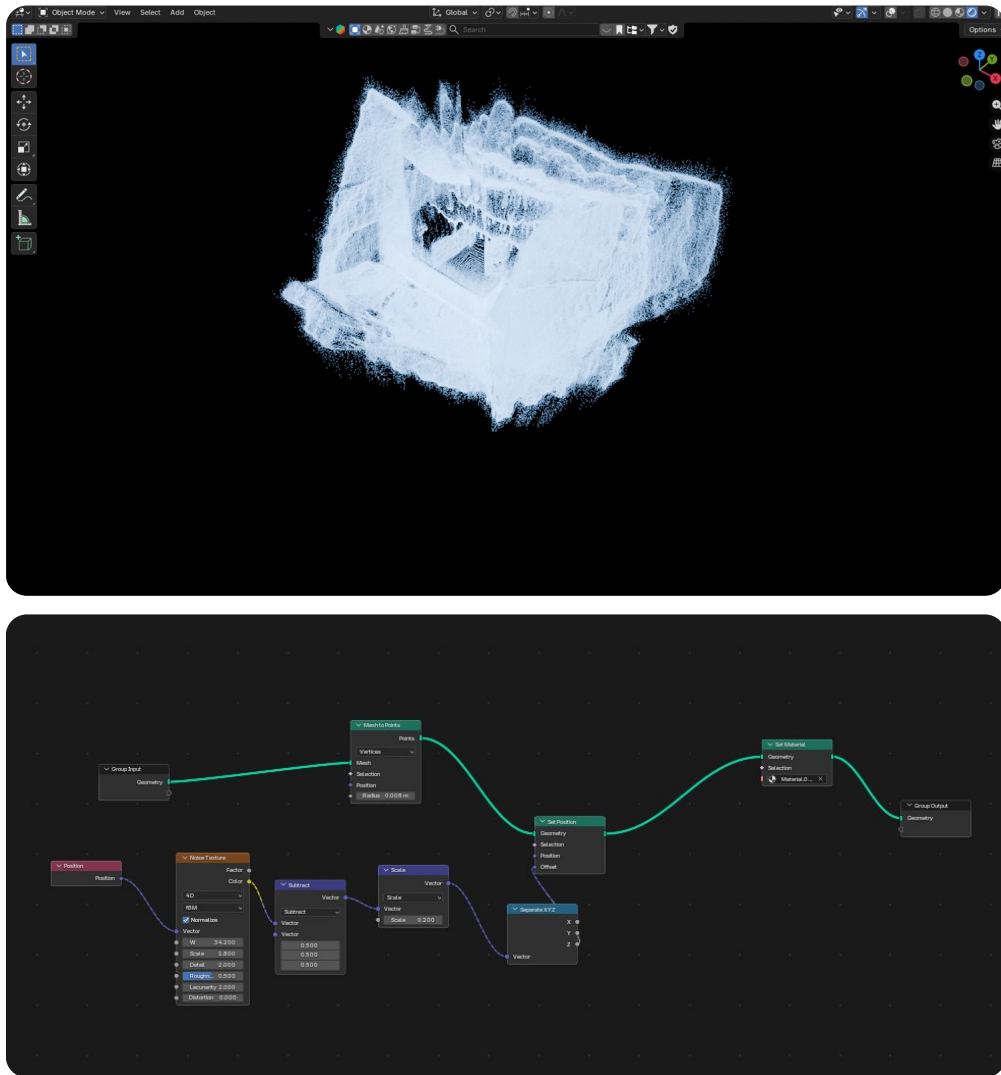


Figura 3.9: Elaborazione della nuvola in Blender: in alto il risultato nel viewport, con il materiale emissivo azzurro; in basso l'albero di *Geometry Nodes* (*Mesh to Points*, *Set Position* pilotato da una *Noise Texture 4D* e *Set Material*).

(rumore procedurale, torsione e onda) conferiva alla nuvola di punti l'aspetto vivo e atmosferico ricercato, pur mantenendo fissa la camera (Figura 3.10).

La pulizia della nuvola è stata effettuata manualmente, selezionando ed eliminando i punti spuri: la ricostruzione risultava comunque piuttosto precisa e l'intervento si è concentrato soprattutto nelle zone in prossimità della camera, dove la densità dei punti era massima. In questa stessa fase sono stati corretti i difetti geometrici evidenziati nel §3.3.2. Il “buco” quadrato che si apriva al di sotto del

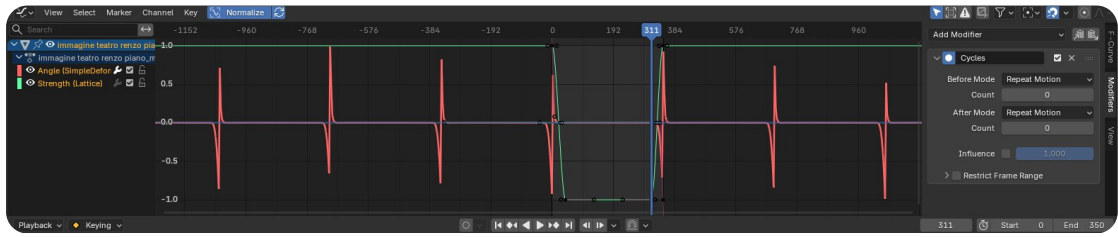


Figura 3.10: Il *Graph Editor* con le curve di animazione dei modificatori (*Angle* del *Simple Deform* e *Strength* del *Lattice*); il modificatore *Cycles* rende l'animazione ciclica sui 350 fotogrammi.

punto di ripresa è stato mascherato applicando alla base della scena un piano dotato dello stesso materiale emissivo dei punti, così da renderlo indistinguibile dal resto della nuvola. L'analoga apertura al polo superiore è stata invece dissimulata da una nebbia di profondità, realizzata mediante due volumi cubici che ne modulavano la densità in modo differenziato (più rada nella parte inferiore e più fitta in quella superiore) così da occultare l'apertura e conferire al contempo maggiore profondità atmosferica alla scena (Figura 3.11).

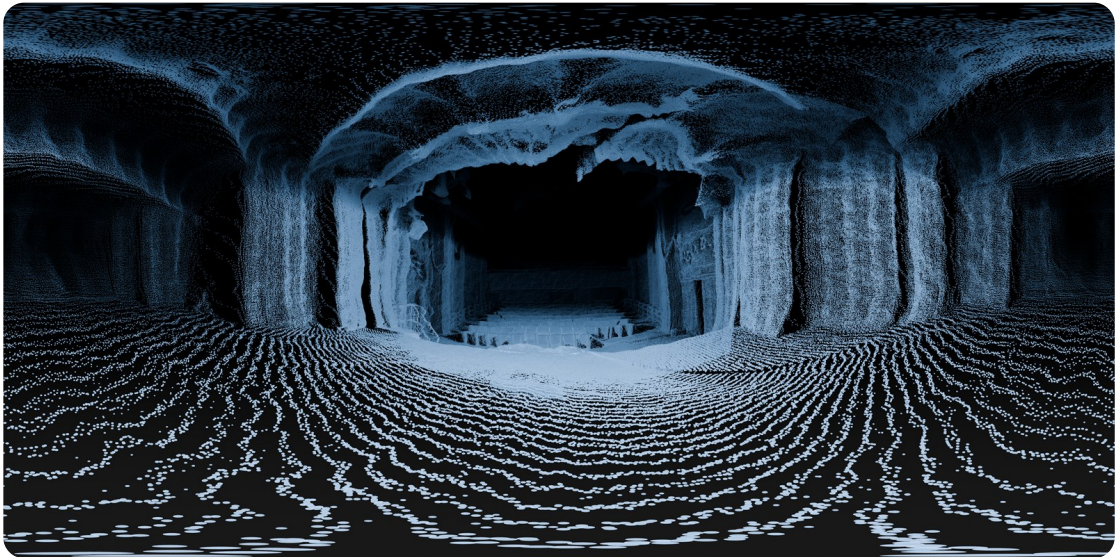


Figura 3.11: Render finale a nuvola di punti di uno degli ambienti ricostruiti (Accademia di Belle Arti dell'Aquila), in proiezione equirettangolare a 360°, con materiale emissivo azzurro e nebbia di profondità.

Il rendering è stato eseguito con il motore Cycles su GPU. Per ottenere l'immagine sferica a 360° la camera adottava una proiezione panoramica equirettangolare (latitudine da -90° a $+90^\circ$ e longitudine da -180° a $+180^\circ$); l'output era prodotto

alla risoluzione di 5760×2880 pixel (nel rapporto 2:1 proprio del formato equirettangolare) con un campionamento contenuto (dieci campioni) affiancato dal *denoiser* OpenImageDenoise, così da mantenere ridotti i tempi di calcolo. L'animazione di ciascun teatro, della durata di 350 fotogrammi, era resa ciclica applicando alle curve di animazione l'apposito modificatore di ripetizione (*Cycles*) dell'editor dei grafici. La finalizzazione del loop e la costruzione delle transizioni tra un ambiente e l'altro sono avvenute in fase di montaggio, con Adobe Premiere: i render dei singoli teatri sono stati assemblati in un unico video equirettangolare a 360° , successivamente impiegato come sfondo animato del menu all'interno di 3DVista (§3.4).

3.4 Architettura e sviluppo dell'interfaccia VR (3DVista)

L'interfaccia di fruizione è stata sviluppata con 3DVista, software di authoring per esperienze immersive e virtual tour interattivi. La scelta di questa piattaforma è stata dettata dalla necessità di garantire un deploy web immediato (tramite standard WebXR): ciò ha permesso di eseguire l'intero progetto direttamente dal browser integrato del visore, senza costringere l'utente all'installazione di un'applicazione dedicata (sideloading). La progettazione della struttura di navigazione e del layout degli elementi interattivi è stata curata dal professor Stefano Sburlati; la realizzazione degli sfondi immersivi a nuvola di punti, la loro integrazione nel progetto 3DVista e la verifica dell'intera esperienza su visore sono state invece condotte dal sottoscritto. L'interfaccia costituisce il punto di accesso all'intera esperienza, raccogliendo in un unico ambiente navigabile i quattro documentari in Realtà Virtuale.

L'idea progettuale è quella di utilizzare le ricostruzioni tridimensionali dei palcoscenici, descritte nel §3.3, come sfondo immersivo dell'interfaccia. A questo scopo è stata realizzata un'unica animazione panoramica in loop, ottenuta concatenando le ricostruzioni a nuvola di punti dei diversi teatri e le relative transizioni. Il video panoramico, renderizzato come contenuto sferico a 360° , costituisce l'ambiente visivo all'interno del quale si svolge l'intera esperienza di selezione, sostituendo

il tradizionale menu bidimensionale con uno spazio immersivo coerente con il linguaggio visivo dei documentari.

Gli elementi interattivi (*hotspot*) sono collocati in posizioni fisse all'interno del panorama e rimangono disponibili durante l'intera animazione. L'interazione avviene mediante il controller del visore, con il quale l'utente seleziona il documentario desiderato; al termine della riproduzione, l'applicazione riporta automaticamente al menu immersivo, permettendo una nuova selezione senza interrompere la continuità dell'esperienza.

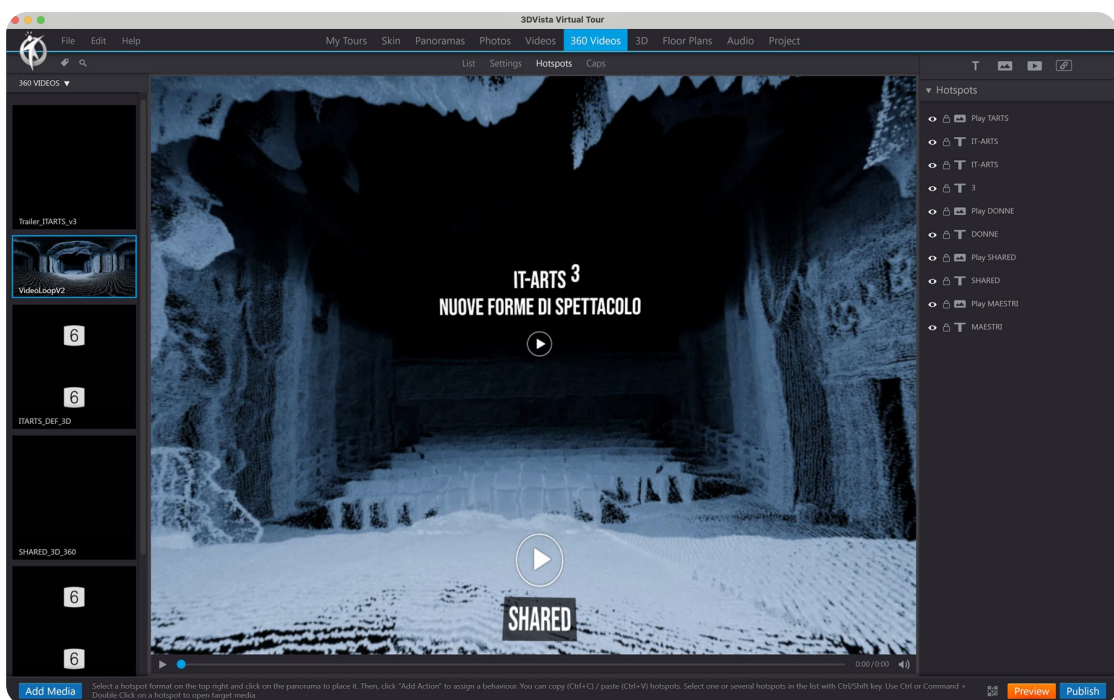


Figura 3.12: Ambiente di authoring in 3DVista: il menu immersivo con lo sfondo animato a nuvola di punti a 360° e gli *hotspot* di selezione dei quattro documentari.

L'intero sistema è stato verificato funzionalmente su Meta Quest 3, sul quale il progetto sviluppato in 3DVista viene eseguito come applicazione web immersiva; in occasione dell'evento conclusivo del progetto, tuttavia, i documentari sono stati fruiti come contenuti singoli, rimandando l'impiego pubblico del menu integrato a una fase successiva (cfr. Capitolo 4). L'interfaccia rappresenta così il punto di convergenza della pipeline descritta nei paragrafi precedenti, integrando in un'unica

esperienza i documentari, le ricostruzioni volumetriche dei teatri e gli strumenti di navigazione.

Capitolo 4

Risultati e Conclusioni

4.1 I deliverable finali della documentazione immersiva

Il prodotto conclusivo del lavoro è costituito da quattro documentari in Realtà Virtuale: tre dedicati alle opere multidisciplinari realizzate per il progetto itARTS (*Donne all’Opera: Croce e Delizia*, *Dei Maestri Dipintori* e *Shared*), cui si affianca un quarto contributo di sintesi dell’intera esperienza. Ciascun documentario è il risultato della pipeline ibrida descritta nel Capitolo 3: le interviste ai protagonisti in VR180 stereoscopico, le riprese a 360° degli spettacoli e l’integrazione dei materiali bidimensionali.

La Tabella 4.1 riassume i quattro deliverable e i materiali specifici integrati, mentre la Tabella 4.2 ne riporta le caratteristiche tecniche.

I quattro file master condividono il medesimo formato di consegna: video **stereoscopici a 360°** in proiezione equirettangolare, con i fotogrammi impacchettati in verticale (*Top-and-Bottom*, Figura 4.1) a una risoluzione di 5760×5760 pixel, codificati in HEVC a 25 fps. Accanto alla fruizione immersiva con visore, i documentari sono stati resi disponibili anche su YouTube, per facilitarne l’accessibilità fuori dal contesto VR dedicato.

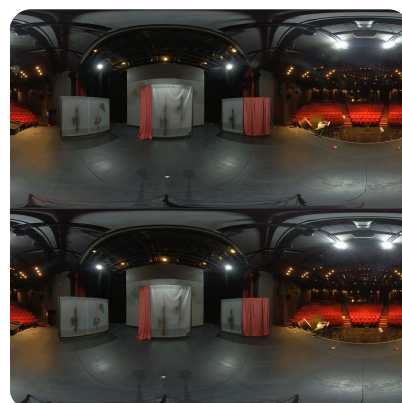


Figura 4.1: Il formato di consegna: fotogramma *Top-and-Bottom* (occhio sinistro sopra, destro sotto), 5760×5760 pixel (*Dei Maestri Dipintori*).

Tabella 4.1: I quattro deliverable del progetto. Ciascun documentario integra le interviste in VR180 stereoscopico e le riprese a 360° di spettacoli e prove; le caratteristiche tecniche di consegna, comuni a tutti, sono riportate nella Tabella 4.2.

Documentario	Focus narrativo	Materiali specifici integrati
<i>Donne all’Opera: Croce e Delizia</i>	Le figure femminili nel contesto operistico e teatrale	Animazioni generate con IA per il LED wall (una per aria); maquette, bozzetti scenografici e studi per i costumi
<i>Dei Maestri Dipintori</i>	Il dialogo visivo e performativo con la tradizione pittorica	Riproduzioni dei dipinti di Paolo Uccello, Giandomenico Tiepolo e Alberto Burri
<i>Shared</i>	La co-creazione e l’interazione tra pubblico e performer	Segni grafici dell’installazione e dell’esperienza <i>SharedVR</i>
Video di presentazione	Introduzione al progetto itARTS e sintesi delle tre produzioni	Materiali di sintesi delle tre opere

Tabella 4.2: Caratteristiche tecniche dei quattro file master, misurate direttamente sui contenitori finali (codifica video HEVC 5760×5760 px, 25 fps; audio AAC stereo, 48 kHz). Le durate sono comprensive dei titoli di coda.

Documentario	Durata	Bitrate video	Dimensione file
<i>Donne all’Opera: Croce e Delizia</i>	10 min 08 s	53,0 Mbps	4,05 GB
<i>Dei Maestri Dipintori</i>	10 min 36 s	51,9 Mbps	4,15 GB
<i>Shared</i>	10 min 04 s	57,0 Mbps	4,33 GB
Video di presentazione	11 min 58 s	48,8 Mbps	4,41 GB

4.2 Discussione dei risultati: sfide tecniche e scelte creative

Il bilancio tecnico della pipeline mette in luce tanto scelte efficaci quanto costi sottostimati. La decisione più efficace si è rivelata l’approccio multi-camera adottato per spettacoli e prove: trattandosi di eventi irripetibili, la ridondanza dei punti di vista ha garantito la copertura di ogni momento significativo e un’effettiva libertà di scelta in fase di montaggio. Al contrario, l’onere della gestione dei dati è

risultato superiore alle previsioni: l'elaborazione di oltre 27 Terabyte di girato ad altissima risoluzione ha assorbito una parte rilevante del tempo di post-produzione, rendendo indispensabili il flusso di lavoro su file proxy e una rigorosa disciplina di duplicazione e archiviazione. Sul fronte della fruizione, i file master HEVC, nonostante l'elevata risoluzione di 5760×5760 pixel, hanno garantito una riproduzione fluida sul visore Meta Quest 3.

Dal punto di vista creativo, il progetto ha esplorato soluzioni per arricchire l'esperienza immersiva. Ad esempio, la sperimentazione legata all'emisfero posteriore delle riprese in VR180 (normalmente lasciato vuoto) ha permesso di inserire bozzetti e visual 2D in post-produzione, con l'intento di densificare la narrazione senza sottrarre attenzione alla scena principale. In parallelo, per la futura navigazione tra i documentari, è stata sviluppata un'interfaccia basata su 3D Gaussian Splatting (generata tramite IA e rifinita in Blender) per creare un ambiente fluttuante e rarefatto, anche se in fase di primo test ci si è concentrati esclusivamente sulla fruizione dei documentari veri e propri.

Una scelta progettuale che merita di essere motivata riguarda l'audio: i master adottano una traccia stereo convenzionale anziché un mix spaziale ambisonico. Si è trattato di una decisione deliberata: data la quantità di ambienti ripresi e la varietà di inquadrature e formati che si alternano all'interno di ciascun documentario, un mix spazializzato coerente avrebbe richiesto di riallineare il campo sonoro a ogni cambio di scena, con una complessità di post-produzione sproporzionata rispetto ai benefici attesi. Si è quindi privilegiata la solidità di uno stereo classico, che garantisce uniformità tra le scene e piena intelligibilità delle voci raccolte dai lavalier, spina dorsale narrativa dei documentari.

L'analisi critica dei deliverable evidenzia invece un limite di natura percettiva: la coesistenza, all'interno dello stesso master stereoscopico, di sorgenti stereoscopiche (le interviste in VR180 e le riprese della Insta360 Pro2) e monoscopiche (le riprese Qoocam, duplicate su entrambi gli occhi): nel passaggio tra le due tipologie la percezione della profondità si attenua bruscamente. L'impatto percettivo è mitigato dal fatto che tali cambi coincidono sempre con uno stacco netto tra ambienti differenti (§3.2.4), ma una copertura interamente stereoscopica rimane un obiettivo per le produzioni future.

4.3 Validazione empirica all'evento itARTS dell'Aquila

La verifica sul campo della documentazione prodotta si è tenuta all'Aquila, in occasione dell'evento conclusivo del progetto itARTS. In tale contesto, il test si è basato sulla fruizione dei video immersivi separati, in attesa della successiva integrazione nell'architettura interattiva in 3DVista.

Oltre trenta persone di estrazione eterogenea hanno provato i visori (Figura 4.2). Le osservazioni sono state raccolte in forma qualitativa: l'assistenza individuale fornita a ciascun utente nell'indossare e avviare il visore ha reso possibile l'osservazione diretta dei comportamenti durante la fruizione, integrata dai commenti spontanei degli spettatori e da brevi domande informali al termine della visione. La scelta di non somministrare questionari strutturati è stata deliberata: trattandosi dell'evento pubblico di chiusura del progetto, con un programma fitto di appuntamenti, trattenere gli spettatori oltre la durata dell'esperienza sarebbe risultato inopportuno e avrebbe sottratto tempo agli altri momenti della giornata; si è quindi privilegiata la spontaneità della fruizione rispetto alla sistematicità della misurazione. Il pubblico ha apprezzato il forte senso di presenza restituito dal formato VR180 durante le interviste e la libertà esplorativa degli spazi a 360°, e la fruizione è risultata confortevole, con una pressoché totale assenza di malessere da simulatore (*motion sickness*).

Un ulteriore elemento emerso con evidenza è l'effetto novità: per la grande maggioranza dei presenti si trattava della prima esperienza in Realtà Virtuale, e una parte consistente dell'esplorazione dello spazio sferico è apparsa guidata più dalla curiosità verso la tecnologia che dal filo della narrazione. Alcuni spettatori sono arrivati a tendere le mani verso i performer, nel tentativo di toccare persone percepite come fisicamente presenti: una manifestazione spontanea, di natura comportamentale, di quella *Place Illusion* descritta nel §1.2.1, per cui l'utente reagisce all'ambiente virtuale in modo analogo a quanto farebbe in un contesto reale [8]. Se da un lato tali reazioni confermano l'intensità del senso di presenza generato dai documentari, dall'altro suggeriscono cautela nell'interpretazione dei riscontri raccolti: parte dell'entusiasmo osservato è verosimilmente attribuibile alla

novità del mezzo più che alle specifiche scelte di linguaggio, un fattore che lo studio strutturato prospettato nel §4.4 dovrà isolare e tenere in considerazione.

È emerso tuttavia un limite legato all'abitudine di fruizione: senza indicatori visivi espliciti, in pochi hanno esplorato autonomamente l'emisfero posteriore notando i contributi 2D alle loro spalle. Questo dato evidenzia come l'attenzione dello spettatore, pur in un ambiente sferico, tenda fisiologicamente a rimanere ancorata all'azione frontale, in linea con quanto documentato dalla letteratura sulla regia immersiva richiamata nel §2.1.3 [16, 34]. È plausibile che la scelta di un audio non spazializzato (§4.2) abbia accentuato questa tendenza, non offrendo allo spettatore lo stimolo sonoro che la letteratura indica come principale mezzo per orientare l'attenzione verso ciò che accade al di fuori del campo visivo [16].



Figura 4.2: Spettatori eterogenei provano i documentari immersivi con il visore Meta Quest 3 durante l'evento conclusivo del progetto itARTS all'Aquila.

4.4 Conclusioni e sviluppi futuri

L'obiettivo di questo elaborato è stato progettare e realizzare la documentazione audiovisiva immersiva per tre opere multidisciplinari promosse all'interno del progetto itARTS. Il lavoro ha dimostrato come l'uso della Realtà Virtuale offra un nuovo paradigma per raccontare lo spettacolo dal vivo: spostando lo spettatore al centro dell'azione, i documentari realizzati permettono di preservare e trasmettere la natura spaziale, relazionale ed effimera della performance, superando i limiti del video bidimensionale tradizionale.

Sul piano metodologico, il risultato principale del lavoro risiede nell'impostazione di un flusso di lavoro ibrido che unisce riprese stereoscopiche, video sferici monoscopici e algoritmi di intelligenza artificiale, fornendo un approccio tecnico documentato che può essere replicato per future produzioni in ambito teatrale e musicale.

Restano tuttavia aperte alcune sfide, che delineano una progressione naturale di sviluppi futuri. Il passo più immediato, già pianificato, è la pubblicazione dell'esperienza integrata sviluppata in 3DVista, che riunirà i quattro documentari nel menu immersivo descritto nel §3.4. L'esperienza di validazione ha confermato l'efficacia del linguaggio adottato, ma uno studio strutturato con gli utenti (tramite questionari standardizzati sul senso di presenza), condotto in un contesto dedicato alla valutazione anziché in un evento pubblico, rappresenterà il naturale passo successivo per quantificare il coinvolgimento emotivo. Infine, dal punto di vista tecnologico, la direzione di sviluppo più promettente è il superamento dell'attuale vincolo dei tre gradi di libertà (3-DoF). La transizione verso un paradigma a sei gradi di libertà (6-DoF), in cui lo spettatore possa muoversi fisicamente all'interno del documentario, richiederà pipeline interamente volumetriche. L'evoluzione di modelli IA capaci di generare scene gaussiane navigabili (come UniSHARP [19]), l'estensione del Gaussian Splatting alle scene dinamiche (4D Gaussian Splatting [36]) e la diffusione di visori sempre più performanti lasciano presagire un'ulteriore evoluzione nelle tecniche di documentazione dell'arte.

Bibliografia

- [1] Paul Milgram e Fumio Kishino. «A Taxonomy of Mixed Reality Visual Displays». In: *IEICE Transactions on Information and Systems* E77-D.12 (1994), pp. 1321–1329 (cit. a p. 2).
- [2] Bruno Rodriguez-Garcia, Henar Guillen-Sanz, David Checa e Andres Bustillo. «A systematic review of virtual 3D reconstructions of Cultural Heritage in immersive Virtual Reality». In: *Multimedia Tools and Applications* 83.42 (2024), pp. 89743–89793. DOI: 10.1007/s11042-024-18700-3 (cit. alle pp. 2, 3).
- [3] Hui Tung Chong, Chen Kim Lim, Ahmad Rafi, Kian Lam Tan e Muslim Mokhtar. «Comprehensive systematic review on virtual reality for cultural heritage practices: coherent taxonomy and motivations». In: *Multimedia Systems* 28.3 (2022), pp. 711–726. DOI: 10.1007/s00530-021-00869-4 (cit. a p. 2).
- [4] UNESCO. *Convention for the Safeguarding of the Intangible Cultural Heritage*. Paris: UNESCO, 2003 (cit. a p. 3).
- [5] Yumeng Hou, Sarah Kenderdine, Davide Picca, Mattia Egloff e Alessandro Adamou. «Digitizing Intangible Cultural Heritage Embodied: State of the Art». In: *ACM Journal on Computing and Cultural Heritage* 15.3 (2022), 55:1–55:20. DOI: 10.1145/3494837 (cit. alle pp. 3, 12).
- [6] Fabio Remondino. «Heritage Recording and 3D Modeling with Photogrammetry and 3D Scanning». In: *Remote Sensing* 3.6 (2011), pp. 1104–1138. DOI: 10.3390/rs3061104 (cit. alle pp. 3, 19).

- [7] Mel Slater e Sylvia Wilbur. «A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments». In: *Presence: Teleoperators and Virtual Environments* 6.6 (1997), pp. 603–616 (cit. a p. 6).
- [8] Mel Slater. «Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments». In: *Philosophical Transactions of the Royal Society B* 364.1535 (2009), pp. 3549–3557. DOI: 10.1098/rstb.2009.0138 (cit. alle pp. 6, 52).
- [9] Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiaogang Xu, Jiashi Feng e Hengshuang Zhao. «Depth Anything V2». In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2024. arXiv: 2406.09414. URL: <https://arxiv.org/abs/2406.09414> (cit. alle pp. 9, 25).
- [10] Lars Mescheder, Wei Dong, Shiwei Li et al. *Sharp Monocular View Synthesis in Less Than a Second*. arXiv preprint. 2025. arXiv: 2512.10685. URL: <https://arxiv.org/abs/2512.10685> (cit. alle pp. 9, 25).
- [11] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler e George Drettakis. «3D Gaussian Splatting for Real-Time Radiance Field Rendering». In: *ACM Transactions on Graphics* 42.4 (2023), pp. 1–14. DOI: 10.1145/3592433 (cit. alle pp. 9, 21).
- [12] Lemonnia Argyriou, Daphne Economou e Vassiliki Bouki. «Virtual Reality with 360-Video Storytelling in Cultural Heritage: Study of Presence, Engagement, and Immersion». In: *Sensors* 20.20 (2020), p. 5851. DOI: 10.3390/s20205851 (cit. alle pp. 12, 15).
- [13] Nonny de la Peña, Peggy Weil, Joan Llobera, Elias Giannopoulos, Ausiàs Pomés, Bernhard Spanlang, Doron Friedman, Maria V. Sanchez-Vives e Mel Slater. «Immersive Journalism: Immersive Virtual Reality for the First-Person Experience of News». In: *Presence: Teleoperators and Virtual Environments* 19.4 (2010), pp. 291–301. DOI: 10.1162/pres_a_00005 (cit. a p. 12).

- [14] Milad Tangshir. *VR Free*. Cortometraggio documentario in Realtà Virtuale (360°), Italia. Presentato alla 76^a Mostra Internazionale d'Arte Cinematografica di Venezia (Venice VR) e al Sundance Film Festival. 2019 (cit. a p. 12).
- [15] F. Tian, Y. Lu, M. Tu et al. «Digital heritage integration of Kunqu opera and Suzhou classical gardens». In: *npj Heritage Science* 14 (2026), p. 78. DOI: 10.1038/s40494-026-02353-6 (cit. a p. 18).
- [16] Lasse T. Nielsen, Matias B. Møller, Sune D. Hartmeyer, Troels C. M. Ljung, Niels C. Nilsson, Rolf Nordahl e Stefania Serafin. «Missing the point: an exploration of how to guide users' attention during cinematic virtual reality». In: *Proc. 22nd ACM Conf. on Virtual Reality Software and Technology (VRST '16)*. 2016. DOI: 10.1145/2993369.2993405 (cit. alle pp. 16, 40, 53).
- [17] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi e Ren Ng. «NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis». In: *Proc. European Conf. on Computer Vision (ECCV)*. 2020. DOI: 10.1007/978-3-030-58452-8_24. arXiv: 2003.08934. URL: <https://arxiv.org/abs/2003.08934> (cit. a p. 20).
- [18] Paolo Clini, Romina Nespeca, Renato Angeloni e Ludovico Coppetta. «3D representation of Architectural Heritage: a comparative analysis of NeRF, Gaussian Splatting, and SfM-MVS reconstructions using low-cost sensors». In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-2/W8-2024* (2024), pp. 93–99. DOI: 10.5194/isprs-archives-XLVIII-2-W8-2024-93-2024 (cit. a p. 23).
- [19] Meixi Song, Dizhe Zhang, Hao Ren, Ruiyang Zhang, Bo Du, Ming-Hsuan Yang e Lu Qi. *UniSHARP: Universal Sharp Monocular View Synthesis*. arXiv preprint. 2026. arXiv: 2606.07514. URL: <https://arxiv.org/abs/2606.07514> (cit. alle pp. 26, 54).
- [20] Enndee. *SHARP_360_to_Splat — Convert stitched 360° equirectangular panoramas into 3D Gaussian Splats*. Repository GitHub. 2026. URL: https://github.com/Enndee/SHARP_360_to_Splat (cit. a p. 26).

- [21] Hualie Jiang, Ziyang Song, Zhiqiang Lou, Rui Xu e Minglang Tan. *Depth Anything in 360°: Towards Scale Invariance in the Wild*. arXiv preprint. 2025. arXiv: 2512.22819. URL: <https://arxiv.org/abs/2512.22819> (cit. a p. 26).
- [22] Carsten Griwodz, Simone Gasparini, Lilian Calvet, Pierre Gurdjos, Fabien Castan, Benoit Maujean, Gregoire De Lillo e Yann Lanthony. «AliceVision Meshroom: An open-source 3D reconstruction pipeline». In: *Proc. 12th ACM Multimedia Systems Conf. (MMSys '21)*. 2021. DOI: 10.1145/3458305.3478443. URL: <https://alicevision.org> (cit. a p. 41).
- [23] Johannes L. Schönberger e Jan-Michael Frahm. «Structure-from-Motion Revisited». In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. COLMAP. 2016 (cit. a p. 41).
- [24] Arthur Brussee. *Brush: 3D Reconstruction for All*. Repository GitHub. URL: <https://github.com/ArthurBrussee/brush> (cit. a p. 41).
- [25] Meta. *Meta Horizon Hyperscape*. Meta Horizon Store (Beta). 2025. URL: <https://www.meta.com/experiences/meta-horizon-hyperscape-preview/7972066712871980/> (cit. a p. 41).
- [26] Kevin Cartelier. *SplataraScan: 3D Scanner & Multiplayer Viewer*. Applicazione per Meta Quest (suite Splatara), SideQuest. 2025. URL: <https://sidequestvr.com/app/55342/splatarascan-3d-scanner-multiplayer-viewer> (cit. a p. 41).
- [27] Gabo Arora e Barry Pousman. *Clouds Over Sidra*. Documentario in Realtà Virtuale (360°) prodotto per le Nazioni Unite in collaborazione con Samsung. Presentato al World Economic Forum di Davos. 2015 (cit. a p. 12).
- [28] Felix & Paul Studios e Cirque du Soleil Média. *Inside the Box of Kurios*. Esperienza in Realtà Virtuale (360° stereoscopico), regia di Félix Lajeunesse, Paul Raphaël e Michel Laprise. Daytime Emmy Award 2016. 2015 (cit. a p. 13).
- [29] Dutch National Ballet e &samhoud media. *Night Fall*. Primo balletto in Realtà Virtuale (360°), coreografia di Peter Leung, musica di Robin Rimbaud. 2016 (cit. a p. 13).

- [30] Kateryna Iudova-Romanova, Tetiana Humenyuk, Serhii Horevalov, Serhii Honcharuk e Volodymyr Mykhalov. «Virtual Reality in Contemporary Theatre». In: *ACM Journal on Computing and Cultural Heritage* 15.4 (2023). DOI: 10.1145/3524024 (cit. a p. 13).
- [31] Google LLC. *VR180 Video Format*. Specifica tecnica del formato, repository Google Spatial Media. 2017. URL: <https://github.com/google/spatial-media/blob/master/docs/vr180.md> (cit. a p. 14).
- [32] Apple Inc. *New Apple Immersive Video series and films premiere on Vision Pro*. Apple Newsroom. 2024. URL: <https://www.apple.com/newsroom/2024/07/new-apple-immersive-video-series-and-films-premiere-on-vision-pro/> (cit. a p. 15).
- [33] John Mateer. «Directing for Cinematic Virtual Reality: how the traditional film director’s craft applies to immersive environments and notions of presence». In: *Journal of Media Practice* 18.1 (2017), pp. 14–25. DOI: 10.1080/14682753.2017.1305838 (cit. alle pp. 16, 39).
- [34] Sylvia Rothe, Daniel Buschek e Heinrich Hußmann. «Guidance in Cinematic Virtual Reality — Taxonomy, Research Status and Challenges». In: *Multimodal Technologies and Interaction* 3.1 (2019), p. 19. DOI: 10.3390/mti3010019 (cit. alle pp. 16, 40, 53).
- [35] Yili Jin, Kaiyuan Hu, Junhua Liu, Fangxin Wang e Xue Liu. «From Capture to Display: A Survey on Volumetric Video». In: *arXiv preprint* (2023). arXiv: 2309.05658. URL: <https://arxiv.org/abs/2309.05658> (cit. a p. 17).
- [36] Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian e Xinggang Wang. «4D Gaussian Splatting for Real-Time Dynamic Scene Rendering». In: *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2024, pp. 20310–20320. arXiv: 2310.08528. URL: <https://arxiv.org/abs/2310.08528> (cit. alle pp. 17, 54).
- [37] Kent Bye e André Lauzon. *Lessons from Cirque du Soleil’s Kurios on Directing Attention in 360 Videos*. Voices of VR Podcast, episodio 141. Intervista ad André Lauzon, produttore di Cirque du Soleil Média. 2015. URL: <https://voicesofvr.com/141-andre-lauzon-on-lessons-from-cirque-du->

soleils-kurios-on-directing-attention-in-360-videos/ (cit. a p. 13).

- [38] MIT Open Documentary Lab. *Interview with Felix & Paul*. Docubase, intervista a Félix Lajeunesse e Paul Raphaël. URL: <https://docubase.mit.edu/lab/interviews/interview-with-felix-paul/> (cit. alle pp. 13, 39).