



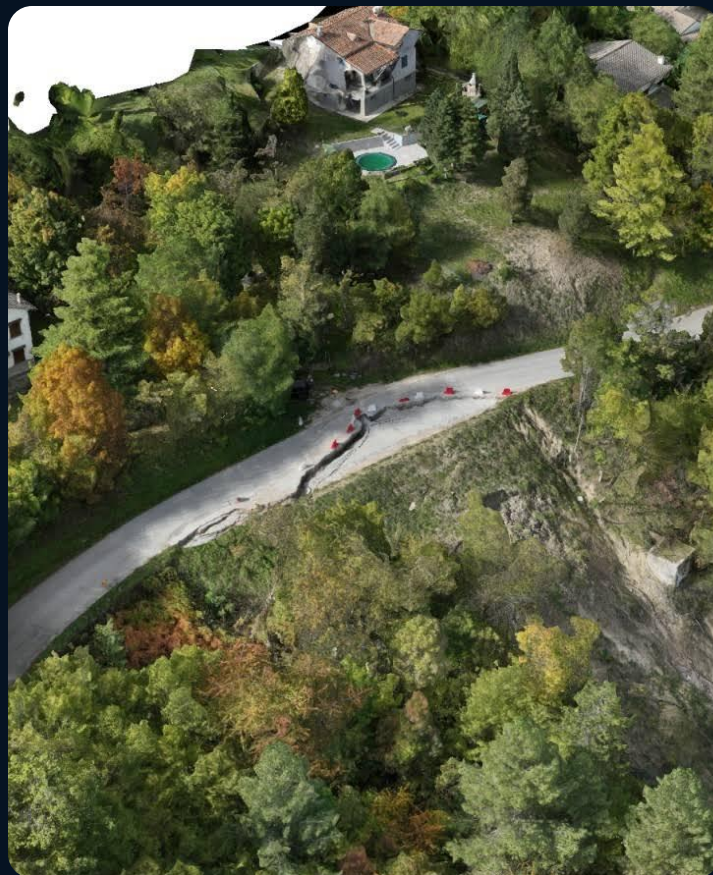
AI, SCIENCE AND HERITAGE DAYS
NAPOLI, 6 OTTOBRE 2026

Dalla fotogrammetria da drone alla Realtà Virtuale

Ricostruzione 3D e navigazione immersiva di scenari colpiti da disastri naturali

Fabio Vitello
INAF - Osservatorio Astrofisico di Catania

Auditorium INAF - Osservatorio Astronomico di Capodimonte



Dove nasce questo lavoro

Il **Competence Center** per la visualizzazione scientifica e la data analytics, all'Osservatorio Astrofisico di Catania.

DAL 2005

Visualizzare l'universo

VisIVO nasce per esplorare grandi dataset astrofisici e simulazioni cosmologiche.

OGGI

Visualizzare il territorio

Analisi geospaziale, modellazione 3D, realtà virtuale e Digital Twin per l'ambiente.

LUGLIO 2026

Un laboratorio

Inaugurato a Catania con il Centro Nazionale HPC, Big Data e Quantum Computing.

Le tecniche sono le stesse: rendere visibile ciò che i dati contengono ma non mostrano.

VISIVOLAB.OACT.INAF.IT

V I S I V  L A B

Dove nasce questo lab

Il **Competence Center** per la
Catania.

DAL 2005

Visualizzare l'universo

VisIVO nasce per esplorare grandi
dataset astrofisici e simulazioni
cosmologiche.

Le tecniche sono le stesse: ren

VISIVOLAB.OACT.INAF.IT



SC

onale di Ricerca in HPC,
and Quantum Computing

DI ASTROFISICA

HAMMON

- Progetto del **Centro Nazionale ICSC** con partner industriali, dedicato a mappare rischi e vulnerabilità legati agli eventi climatici estremi
- **Rilievi UAV e fotogrammetria** per modelli 3D ad alta risoluzione delle aree colpite
- **Applicazione** per l'ispezione delle aree danneggiate
- Tecniche di **AI: segmentazione**; valutazione automatica dei danni **in sviluppo**

Un caso reale: Tredozio (FC)

Alluvioni e frane, maggio 2023

terremoto Mw 4.9, 18 settembre 2023

Rilievo UAV HAMMON, 14–16 ottobre 2024

INAF-OACT con UniMiB ed ENEA



Le domande che guidano il lavoro

Quali edifici sono danneggiati, e quanto?

Coperture, crolli, lesioni visibili.

Come è cambiato il terreno?

Frane, erosioni, variazioni topografiche.

Dove intervenire prima, in sicurezza?

Priorità e percorsi in sicurezza.

Come documentare lo stato dei luoghi?

Per la ricostruzione, la memoria, il patrimonio.

Il percorso di oggi

1

**Dalle immagini al
modello 3D**

2

**Dal modello
all'informazione**

3

Dentro lo scenario

4

Verso il Digital Twin

1

Dalle immagini al modello 3D

Rilievo UAV e fotogrammetria

Da dove viene questo metodo



Il nostro lavoro sta nell'ultimo riquadro: la novità non è ricostruire, ma **ricostruire per decidere** e poterci "entrare dentro".

Perché da più fotografie nasce la terza dimensione

IL PRINCIPIO: LA PARALLASSE

Lo spostamento apparente di un oggetto fra due scatti dipende dalla sua distanza: se sappiamo dove stavano le camere, diventa una misura.



Come i nostri occhi: la visione stereoscopica umana sfrutta esattamente questo principio per percepire la profondità.



Perché da più fotografie nasce la terza dimensione



IL PROBLEMA E LA SOLUZIONE

Un drone non conosce con precisione la propria posa a ogni scatto. Ricavare l'orientamento perfetto delle immagini è la vera sfida della fotogrammetria moderna.

Structure from Motion: rovescia il problema ricavando *insieme* posizioni delle camere e forma 3D della scena dalle sole immagini.

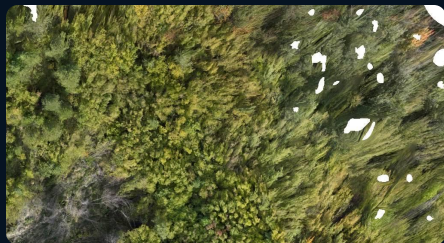
Quando funziona, e quando no

✓ CONDIZIONI OTTIMALI

- **Texture:** superfici con dettaglio riconoscibile
- **Sovrapposizione:** ogni punto presente in più immagini
- **Illuminazione stabile:** luce uniforme senza intermittenze
- **Scena ferma:** geometria invariata durante il volo
- **Immagini nitide:** assenza di sfocature da movimento

⚠ QUI LA RICOSTRUZIONE SOFFRE

- **Acqua:** varia a ogni scatto, priva di punti fissi
- **Vetro e riflessi:** riflettono prospettive differenti
- **Vegetazione nel vento:** si muove tra le riprese
- **Superfici omogenee:** neve, sabbia, muri uniformi
- **Ombre nette:** creano dettagli che si spostano col sole



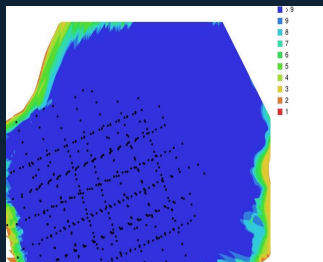
CASO REALE - MONTE BUSCA

Il bosco rappresenta il punto critico: **ottimo dettaglio sulla strada**, ma ricostruzione incerta sotto le chiome. In emergenza le condizioni ideali sono rare: è fondamentale saper valutare quali porzioni del modello 3D siano realmente affidabili.

Pianificare il volo: da cosa dipende il dettaglio

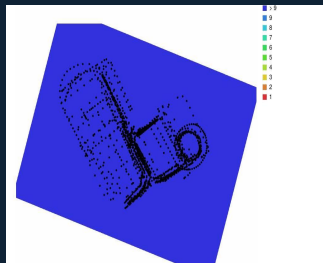
$$\text{GSD} = \text{quota} \times \text{dimensione del pixel} \div \text{focale}$$

Stesso comune, stessa settimana: **versante** (39,6 m → 1,17 cm/pixel) · **piazza e campanile** (7,6 m → 4,2 mm/pixel)



Volo sul versante

Strisciate parallele ad alta quota; copertura ad ampia scala del territorio.



Volo sulla piazza

Percorsi orbitali e ravvicinati attorno agli edifici ed elementi verticali.

Quota di volo

Più bassa = più dettaglio, ma richiede più strisciate, più batterie e aumenta il rischio di ostacoli.

Sovrapposizione (80–85%)

Fiduciale per la ricostruzione 3D: ogni punto del terreno deve comparire in più scatti consecutivi.

Geometria delle traiettorie

Strisciate parallele per rilievi del terreno; orbite e inclinazioni diverse attorno agli edifici.

Dalle fotografie alla geometria

1

Punti in comune

Gli stessi dettagli vengono riconosciuti in immagini diverse (feature matching).

2

Posizioni di scatto

La Structure from Motion ricava dove si trovava la camera per ogni foto.

3

Nuvola densa

Il Multi-View Stereo ricostruisce milioni di punti nello spazio 3D.

4

Georeferenziazione

Le coordinate delle camere e gli eventuali punti di controllo riferiscono il modello al territorio.

Centinaia o migliaia di immagini diventano un unico spazio 3D coerente e misurabile.

Allineamento: dove stavano le camere

1 Punti caratteristici

In ogni immagine si cercano dettagli riconoscibili da angolazioni diverse.

2 Corrispondenze

Lo stesso punto ritrovato in più scatti: sulla piazza 3,1 milioni di punti legati, con oltre 14 milioni di proiezioni.

3 Bundle adjustment

Si risolvono insieme posizioni delle camere, parametri ottici e coordinate dei punti, minimizzando l'errore di riproiezione.



I PARAMETRI STIMATI

f · distanza focale cx, cy · centro ottico

$k1, k2, k3$ · distorsione radiale $p1, p2$ · distorsione tangenziale

Senza questa correzione le rette non tornano rette e il modello si incurva.

Risultato sulla piazza: 2.087 immagini allineate su 2.099, ogni punto visto in media da 5 scatti.

Dalle mappe di profondità alla nuvola densa



1 MAPPE DI PROFONDITÀ

Per ogni immagine si stima la distanza di ogni pixel, confrontandola con le vicine.

Sulla piazza: 2.087 mappe.

2 FUSIONE

Le mappe si combinano in un'unica nuvola:

142 M di punti
con colore e affidabilità.

3 SUPERFICIE

Dai punti si costruisce la mesh geometrica:

51 M di triangoli
texturizzata a
8192x8192.

Le due manopole che contano

Qualità: a quale scala si lavorano le immagini, ogni passo raddoppia il dettaglio e quadruplica il tempo.

Filtro di profondità: quanto aggressivamente si scartano i punti incerti.

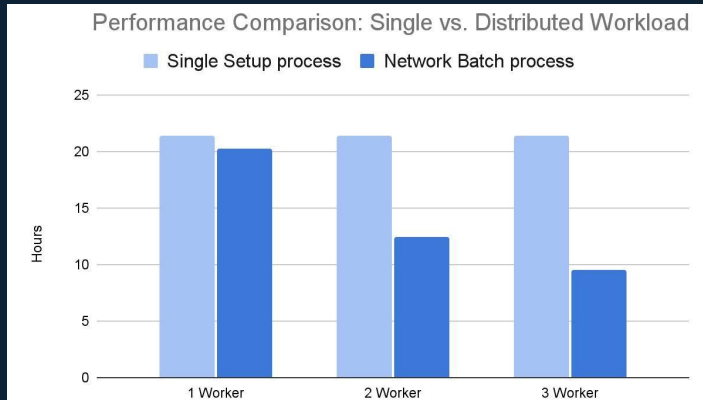
Piazza

Qualità alta, filtro moderato.

Il prezzo: circa **21 ore** di elaborazione complessiva, di cui quasi 7 per la sola nuvola densa.

HPC4DR - perché serve il supercalcolo

BENCHMARK DI CALCOLO



21h → 9,5h
-55% tempo di calcolo

3 Worker
2.099 immagini

1 PERFORMANCE MISURATA

Il risultato misurato

Rilievo della piazza (2.099 immagini): da oltre **21 ore** su singolo nodo a circa **9,5 ore** distribuendo su tre worker.

2 RETE DI COLLABORAZIONE

Il consorzio

Università di Abruzzo, Marche e Molise con CNR, INAF, INFN e INGV; infrastruttura ai Laboratori Nazionali del Gran Sasso.

3 RUOLO OPERATIVO

Il contributo INAF

Portare la pipeline sull'infrastruttura e applicare algoritmi di Machine Learning a danni e cambiamenti.

Tre prodotti, tre usi



Nuvola di punti

misurare e confrontare



Mesh

analizzare la superficie

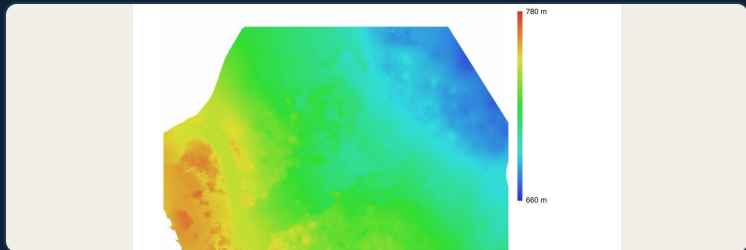


Modello texturizzato

esplorare in fotorealismo

Dagli stessi dati derivano anche ortofoto e modelli digitali del terreno (DEM).

I prodotti per chi lavora in GIS



Modello di elevazione (DEM)

Una quota per ogni cella: la forma del terreno.

Versante di Monte Busca: 11.243×9.598 celle a 2,33 cm

Serve per: pendenze, volumi, profili, confronti fra rilievi



Ortomosaico

Immagini raddrizzate e cucite: una mappa misurabile.

Versante di Monte Busca: 22.483×19.193 pixel a 1,17 cm

Serve per: cartografia, perimetrazioni, documentazione

Sono **2,5D**: una quota per posizione. Non descrivono strapiombi, sottosquadri e facciate; per quelli serve il modello 3D.

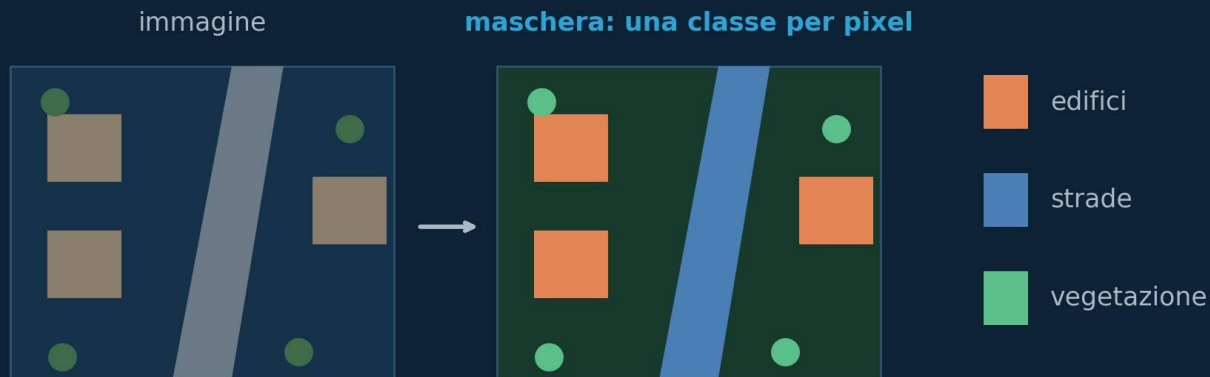
Dalle fotografie al modello



2

Dal modello all'informazione

Segmentazione semantica: dare un nome a ogni pixel



01

Quali pixel, non solo cosa

è questo che permette di riportare il risultato sulla geometria.

02

Serve un maestro

Il modello impara da migliaia di immagini già annotate a mano, non da regole scritte da noi.

03

Le classi sono una scelta

Decidere quali categorie usare è già una decisione di progetto: ciò che non è previsto non verrà visto.

Dati, annotazione, valutazione

1 DATI

Da dove vengono gli esempi

Dataset pubblici di immagini da drone su aree colpite

RescueNet e **FloodNet** più le nostre immagini annotate a mano.

2 ANNOTAZIONE

Annotare costa

È il collo di bottiglia vero. Usiamo **Label Studio** con **SAM**, che propone il contorno a partire da un clic: si corregge invece di disegnare.

3 VALUTAZIONE

Come si misura il risultato

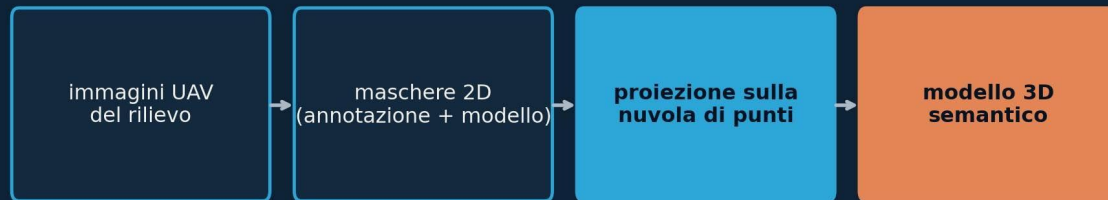
Si confronta ciò che il modello prevede con ciò che una persona ha annotato: è l'unico modo di dire se funziona.



$$\text{IoU} = \frac{\text{sovrapposizione}}{\text{unione}}$$

1 = perfetto, 0 = nessuna corrispondenza

Dalle maschere 2D al modello 3D semantico



la conoscenza nasce in 2D e viene trasferita alla geometria 3D

01

Il trucco sta nella fotogrammetria

Sappiamo già quale punto 3D corrisponde a quale pixel di ogni immagine: è il risultato dell'allineamento. La classe viaggia lungo quel legame.

02

In pratica

La rete produce maschere in bianco e nero per ciascuna classe; Metashape le usa per classificare i punti corrispondenti del modello.

03

Perché in 3D

Le classi permettono di isolare: confrontare solo il terreno, escludere la vegetazione che cresce da sola. Senza, il cambiamento vero si confonde con il rumore.

Che cosa l'AI non fa (ancora)

1 GENERALIZZAZIONE

Non generalizza da sola

Un modello addestrato su alluvioni texane non conosce i borghi appenninici: tetti, materiali, vegetazione e luce sono diversi. Va riaddestrato sul contesto.

2 EXPLAINABILITY

Non spiega da sola

Restituisce una classe, non una motivazione. Per un tecnico che firma un'ordinanza serve poter vedere *perché*: è il lavoro sulla explainability.

3 RESPONSABILITÀ

Non decide

Segnala dove guardare e riduce il lavoro manuale. La valutazione di agibilità resta di chi ha competenza e responsabilità.

Stato del progetto - Caso studio

Sul nostro caso: la **segmentazione è stata eseguita** e i risultati sono preliminari; valutazione del danno e confronto fra rilievi sono **in sviluppo**.

3

Dentro lo scenario

Unity e Realtà Virtuale

Che cos'è, davvero, la realtà virtuale

«L'esperienza in cui l'utente è effettivamente immerso in un mondo virtuale che risponde.»

— Fred Brooks, 1999

PERCEZIONE VISIVA E SPAZIALE

Immersione

La sensazione di essere dentro: il mondo circonda l'osservatore invece di stare in una finestra.

INTERAZIONE IN TEMPO REALE

Responsività

L'ambiente reagisce in tempo reale ai movimenti e alle azioni: senza questo è un film a 360 gradi.

Insieme, i due elementi fanno della VR non un'esperienza visiva, ma una forma di interazione con lo spazio.

Un'idea vecchia di sessant'anni



INNOVAZIONE STORICA

Il Sensorama, 1962

Morton Heilig brevetta una cabina con film stereoscopico, suono stereo, vibrazioni, vento e odori.

OSTACOLI TECNOLOGICI

Che cosa mancava

Non l'idea, ma la tecnologia: potenza di calcolo, schermi, sensori di movimento.

Tre livelli di immersione

1 DESKTOP & MOUSE

Non immersiva

Il mondo virtuale si guarda su uno schermo normale e si controlla con mouse e tastiera. È la forma più diffusa ed economica, e non va sottovalutata.

2 SCHERMI & PROIEZIONI

Semi-immersiva

Grandi schermi, proiezioni o più monitor circondano l'osservatore, che però resta consapevole della stanza in cui si trova.

3 VISORE HMD

Immersiva

Il visore sostituisce completamente il campo visivo. È quella di cui parliamo oggi.

Non esiste il livello giusto in assoluto: dipende da quante persone devono guardare insieme, da quanto deve durare la sessione e da cosa si deve fare.

Tre famiglie di visori

Entry Level

Con smartphone

Un guscio con due lenti in cui si inserisce il telefono. Costa pochissimo, ma la qualità è limitata e manca il tracciamento della posizione: ci si guarda intorno restando fermi.

Standalone

Autonomi

Dispositivi tutto-in-uno, senza fili, con elaborazione integrata. Comodi e pratici, con la potenza di un tablet.

High Performance

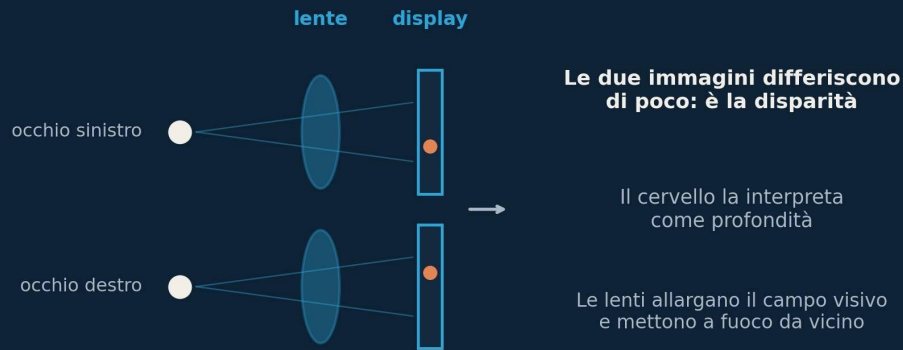
Collegati al PC

Display dedicati, tracciamento preciso, e a disposizione la scheda grafica del computer. La soluzione più costosa e la più performante.

QUALE SCEGLIERE

Per modelli da decine di milioni di triangoli serve la **terza categoria**. Per la divulgazione su scene alleggerite, la **seconda** può essere sufficiente.

Che cosa fa un visore



01 Stereoscopia

La scena viene disegnata due volte, da due punti di vista distanti quanto i nostri occhi.

02 Ottica

Le lenti allargano il campo visivo; l'immagine è pre-deformata per compensarle.

03 Tracciamento

Camere sul visore osservano l'ambiente e ne ricavano la posa: la stessa idea della fotogrammetria, in tempo reale.

Perché un game engine

1 ESPLORAZIONE

Tempo reale

Il modello si esplora, non solo si osserva.

2 MISURAZIONE

Interazione

Misure e annotazioni costruite sulla scena.

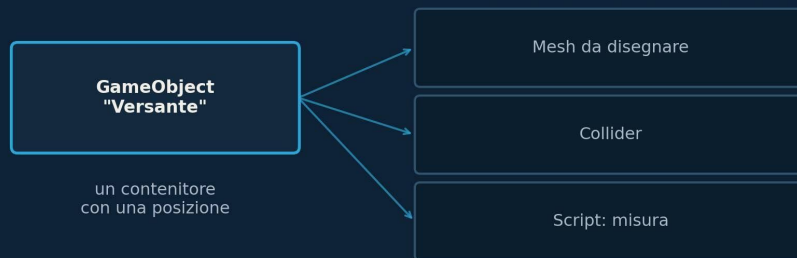
3 MULTI-PIATTAFORMA

Ecosistema XR

L'engine copre desktop, web e visori;
la nostra applicazione: desktop e VR.

Da qui seguiamo il modello di Tredozio dentro [GeaVR](#), il nostro pacchetto open source, fino al visore.

Anatomia di una scena



1 STRUTTURA

GameObject

Un contenitore vuoto con una posizione. Il terreno, la camera, un waypoint: tutto è un GameObject.

2 COMPORTAMENTO

Component

Ciò che gli dà comportamento: mesh da disegnare, collider, sorgente audio, script.

3 RIUSABILITÀ

Prefab

Un oggetto già montato, riusabile: il marcatore che creiamo una volta e istanziamo cento volte.

FLUSSO DI ESECUZIONE

Il ciclo di vita

Uno script si aggancia a questi momenti: **Start** una volta all'avvio, **Update** a ogni fotogramma, **FixedUpdate** a intervalli regolari per la fisica.

CASO D'USO

Nel nostro caso

Il modello del versante è un GameObject con mesh e collider; gli strumenti di misura sono script che leggono i controller.

Il modello entra in Unity

ESPORTAZIONE

Il modello esce a tile: geometria OBJ e texture da 4096×4096 pixel, georeferenziate.

Problema: troppo pesante per il tempo reale

Il convertitore serializza le mesh e converte le texture in DDS, mantenendole compresse nella memoria GPU.

Problema: scala e coordinate

Distinte: la **scala metrica** (1 unità = 1 metro) dà dimensioni reali; la **georeferenziazione** dà coordinate ai dati raccolti.

IL FLUSSO DI CONVERSIONE

Metashape

tile OBJ + texture 4096×4096, WGS84



Convertitore GeaVR

mesh serializzate (SOBJ) + texture DDS



Scena Unity

caricamento a runtime, scala metrica

Il vincolo che decide tutto: il ritardo



sotto i 20 ms, o il movimento si sente in ritardo

1 Perché conta

Se muovo la testa e la scena arriva in ritardo, il mondo "scivola": è la prima causa di malessere.

2 Previsione

Il sistema stima dove sarà la testa quando l'immagine sarà pronta, e disegna quella posizione.

3 Correzione finale

Appena prima di mostrarla, l'immagine viene riallineata all'ultima posa misurata.

RIFERIMENTO: soglia di ~20 ms indicata dalla letteratura e dalla pratica industriale

Il budget di un fotogramma

tutto deve stare qui dentro, 90 volte al secondo



COSTO DI DISEGNO

Draw call

Ogni comando di disegno costa.
Molti oggetti piccoli e separati
costano più di pochi oggetti grandi.

OTTIMIZZAZIONE

Batching

Oggetti che condividono lo stesso
materiale possono essere disegnati
insieme: meno comandi.

DIAGNOSTICA

Profiler

Prima di ottimizzare, misurare: dice
se il collo di bottiglia è la CPU o la
GPU.

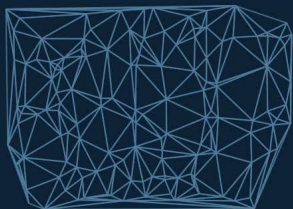
Renderlo fluido

PIRAMIDE DI DETTAGLIO

Gestione LOD

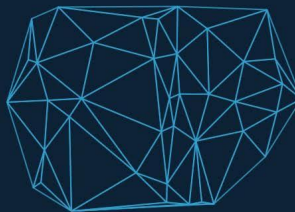
Riduzione dinamica della complessità geometrica in base alla distanza dalla camera.

modello originale



25,2 milioni di triangoli

decimato



geometria alleggerita

a distanza (LOD)



ancora meno dettaglio

APPLICATO AL NOSTRO MODELLO

Decimazione mirata

Passaggio guidato dai **51 milioni di triangoli** del modello originale della piazza a una scena perfettamente navigabile ed efficiente.

ALTRE LEVE DISPONIBILI

- Atlas & compressione texture
- Culling (Frustum / Occlusion)
- Illuminazione precalcolata

OBIETTIVO VR

72-90

FPS (fotogrammi/sec)

Ritmo fondamentale in VR: si ottimizza per mantenere questa fluidità (varia col visore).

Renderlo esplorabile

FISICA & COLLISIONI

Collider

Camminare sul terreno senza attraversarlo.

INTERAZIONE & INPUT

Controller

Movimento, camera e input dell'utente.

INTERFACCIA UTENTE

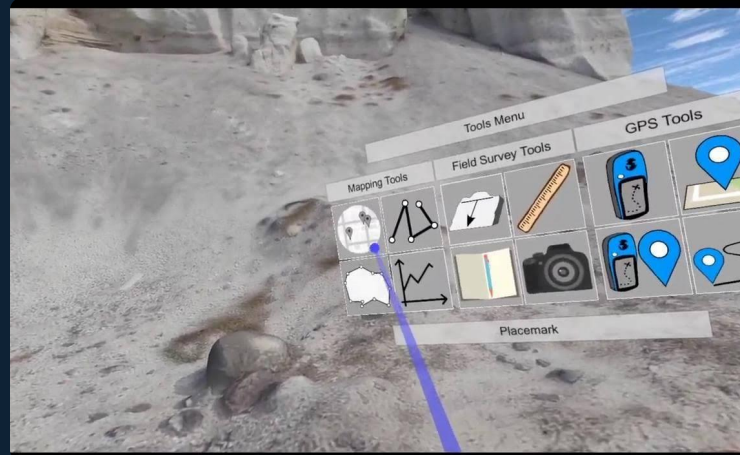
Interfaccia

Menu, minimappa, informazioni sui luoghi.

DATI ANCORATI

Punti di interesse

Foto, dati e analisi ancorati ai luoghi.



L'INTERFACCIA DI GEAVR IN USO SU UNO SCENARIO DI TEST

Portarlo nel visore

SOFTWARE OPEN SOURCE

GeaVR

Pacchetto Unity per esplorazione immersiva e dati
<https://github.com/GeaVR/GeaVR>

HARDWARE TESTATO

Visori Consumer

Oculus Rift, Rift S e Meta Quest 2 collegati a PC da gaming

REQUISITI HARDWARE

Configurazione PC

CPU i5, 16 GB di RAM e GPU VR-ready con 6 GB VRAM

SVILUPPI FUTURI

NavionVR

Evoluzione del pacchetto software in fase di sviluppo



Navigare senza disorientare

Sessioni brevi

Su **459 partecipanti**, sessioni di **10'**: 26% disturbi transitori, **nessuna interruzione** (Bonali et al., 2022).

Comfort

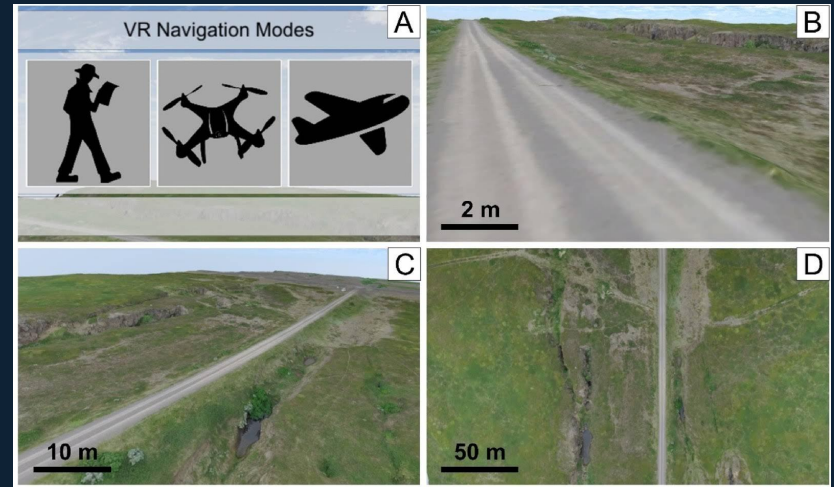
Prestazioni stabili e navigazione adeguata.

Scala metrica reale

1 unità = 1 metro: dimensioni corrispondenti al reale.

Gli accorgimenti che aiutano: rotazione a scatti, riduzione del campo visivo durante lo spostamento, accelerazioni dolci, un riferimento fisso nella scena

Modalità di navigazione VR



Strumenti dentro lo scenario

Mappatura

Punti, linee, profili — export CSV/KML

• Disponibile

Field survey

Righello, giaciture, taccuino

• Disponibile

Strumenti GPS

Posizione, mappa, waypoint

• Disponibile

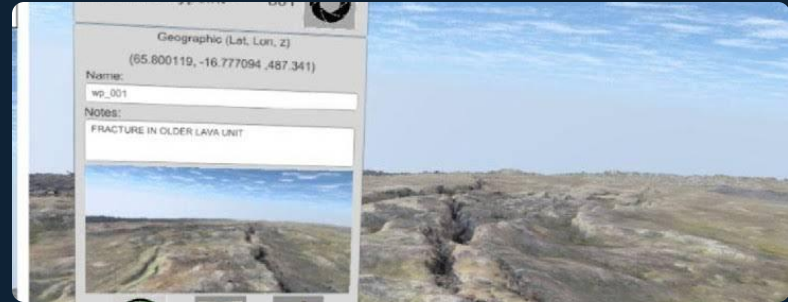
Prima / dopo

Rilievi sovrapposti nel tempo

○ Previsto



Misurare: lunghezza, dislivello, direzione.



Annotare: waypoint con coordinate, nota e foto.



<https://youtu.be/lfqusHa5xIk>



<https://www.youtube.com/watch?v=dUIKDC3HD10>



https://www.youtube.com/watch?v=_yqTdhm0-xY



https://youtu.be/_CFdRXJvuF

4

Dal modello statico al Digital Twin

La prospettiva ECHO-TWIN

Verso l'aggiornamento nel tempo del modello

ECHO-TWIN (progetto PN-RIC coordinato da ICSC, avviato a maggio 2026) punta a trasformare la pipeline post-evento in un ciclo di monitoraggio continuo.

Due scale di volo

Droni leggeri e manovrabili per facciate e spazi confinati; droni pesanti con LiDAR per le grandi aree in erosione.



Riconoscere le lesioni sul posto

Reti neurali ultraleggere che individuano le crepe direttamente sul flusso video, mentre il drone vola.



Laboratorio UAV-Edge e VR

Il Laboratorio di Visualizzazione Scientifica INAF si potenzia: navigazione 3D/VR dei gemelli digitali.

Tre cose da portare a casa

- 1 Un rilievo da drone diventa un **modello 3D misurabile** del territorio.
- 2 AI e HPC lavorano su quel modello: **segmentazione**, analisi di danni ed evoluzioni temporali.
- 3 La VR rende lo scenario **esplorabile oggi**; il Digital Twin punta a renderlo **aggiornabile nel tempo**.

Riferimenti

FOTOGRAMMETRIA

Westoby et al. (2012), *Structure-from-Motion photogrammetry: a low-cost, effective tool for geoscience applications*. *Geomorphology* 179, 300–314

James & Robson (2014), *Mitigating systematic error in topographic models derived from UAV and ground-based image networks*. *ESPL* 39, 1413–1420 · accesso aperto

Furukawa & Hernández (2015), *Multi-View Stereo: A Tutorial*. *Found. Trends Comput. Graph. Vis.* 9(1–2)

Triggs et al. (2000), *Bundle Adjustment — A Modern Synthesis*. *LNCS* 1883, 298–375

DANNI, AI E REALTÀ VIRTUALE

Kerle et al. (2020), *UAV-Based Structural Damage Mapping: A Review*. *ISPRS Int. J. Geo-Inf.* 9(1), 14 · accesso aperto

Rahnemoonfar et al. (2021), *FloodNet*. *IEEE Access* 9, 89644–89654 · dataset pubblico

Jégou et al. (2017), *The One Hundred Layers Tiramisu* — la rete usata per la segmentazione

Kirillov et al. (2023), *Segment Anything*. *ICCV*, 4015–4026 — usato per accelerare l'annotazione

Krokos et al. (2019), *Workflows for virtual reality visualisation and navigation scenarios in Earth Sciences*

Tibaldi, Bonali, Vitello et al. (2020), *Real world–based immersive Virtual Reality for research, teaching and communication in volcanology*. *Bull. Volcanol.* 82, 38

Bonali, Vitello et al. (2024), *GeaVR: an open-source tools package for geological-structural exploration and data collection using immersive virtual reality*. *Applied Computing and Geosciences* 21, 100156 · accesso aperto

Software: geavr.eu (GPL-3.0) · codice HAMMON: github.com/ICSC-Spoke3 - pipeline UAV e machine learning

V I S I V  L A B

Grazie

Domande?

Fabio Vitello

INAF – Osservatorio Astrofisico di Catania

fabio.vitello@inaf.it