SplatTwin: ricostruzione genAI georiferita di aree geografiche da voli drone

Emanuele Flebus[1], Giovanni Lughi[1], Stefano Olivucci[2], Marco Pesic[1], Matteo Roffilli[1]
[1] Bioretics srl, [e.flebus, m.pesic, g.lughi, m.roffilli]@bioretics.com
[2] Regione Emilia-Romagna, Stefano.Olivucci@regione.emilia-romagna.it

Abstract. La ricostruzione digitale fotorealistica e numericamente adeguata a scopi professionali di ampie aree geografiche viene solitamente ottenuta da voli droni o aerei dotati di tecnologie di scansioni laser come il LiDar. Quale estensione di un precedente lavoro presentato sempre ad Asita [1] proponiamo l'utilizzo di tecniche innovative di Intelligenza Artificiale Generativa quali NeRF, 2DGS, e 3DGS, per la creazione di digital twin di strutture quali ponti, argini, elettrodotti, impianti fotovoltaici, ospedali, fino a porzioni di territori o città, a partire unicamente da un insieme di fotografie o video, anche quando acquisiti con tecnologie amatoriali. La pipeline proposta, denominata SplatTwin, georeferenzia automaticamente i modelli a partire dalla telemetria del drone e li pubblica come 3D Tiles consultabili e misurabili via browser, mantenendo il dato sorgente all'interno dell'infrastruttura dell'ente.

Parole chiave: digital-twin, gemello digitale, volo drone, 3D Gaussian Splatting , Cesium, territorio.

1. Introduzione

La disponibilità di gemelli digitali tridimensionali, fotorealistici e metricamente affidabili del territorio è ormai un requisito ricorrente per la pianificazione, il monitoraggio delle infrastrutture e la gestione delle emergenze. Le tecniche consolidate di rilievo LiDAR da aeromobile o drone garantiscono accuratezze elevate ma richiedono sensori dedicati, personale specializzato e tempi di elaborazione che si misurano in giorni. Le tecniche di Intelligenza Artificiale Generativa (GenAI) per la ricostruzione di campi di radianza, dai NeRF [2,3] fino al 2D e 3D Gaussian Splatting [4,5], hanno dimostrato che una ricostruzione fotorealistica e geometricamente coerente è ottenibile a partire da un semplice video, anche amatoriale. Quale estensione del lavoro presentato ad ASITA 2024, questo contributo presenta SplatTwin, una pipeline end-to-end che da un singolo video di volo drone con traccia GPS accessoria produce in pochi minuti un modello 3D in scala metrica assoluta, correttamente posizionato e orientato sul globo, e lo pubblica automaticamente in formati standard per la fruizione via browser e l'integrazione nei GIS. Quando priva dei dati GPS la pipeline produce

comunque un modello geometricamente corretto ma senza il dimensionamento e posizionamento assoluto.

Rispetto al lavoro precedente, i contributi principali sono quattro: (i) la ricostruzione è affidata a un'implementazione del 3D Gaussian Splatting basata sulla libreria open-source gsplat [8], che a parità di qualità finale riduce sensibilmente i tempi di addestramento e consente il rendering in tempo reale; (ii) la georeferenziazione dei modelli è completamente automatica, a partire dalla sola telemetria registrata dal drone; (iii) la pubblicazione come 3D Tiles [7] su globo virtuale è integrata nella pipeline; (iv) l'approccio è validato su casi reali eterogenei: monumenti urbani, versanti montani e aree alluvionate. Il caso d'uso che meglio ne illustra l'utilità è proprio quello emergenziale: la Figura 1 mostra il modello georiferito della rottura dell'argine di Traversara ricostruito dal solo video di sorvolo fornito dalla Regione Emilia-Romagna, quando la tempestività della valutazione può fare la differenza.



Fig. 1. Modello 3D point cloud della rottura dell'argine di Traversara (Bagnacavallo, Ravenna, 2024) sul fiume Montone da video di volo drone fornito dalla Regione Emilia-Romagna.

2. La pipeline SplatTwin

La pipeline di elaborazione è articolata in stadi eseguiti in sequenza senza intervento manuale: (1) estrazione e preprocessamento dei fotogrammi dal video; (2) stima delle pose di camera e di una nuvola sparsa mediante Structure-from-Motion (COLMAP) con bundle adjustment; (3) addestramento del modello neurale di scena, mediante backend NeRF accelerato con hash encoding o 3D Gaussian Splatting implementato con gsplat [8]; (4) esportazione di una nuvola di punti densa (o degli splat stessi) e dei rendering; (5) georeferenziazione automatica mediante traccia GPS (Sezione 3); (6) pubblicazione come file semplice o online in 3D Tiles (Sezione 4). I modelli sono esportabili nei comuni formati già in uso per utilizzi tecnici e di video-produzione:

nuvole di punti .las\.ply, mesh, video di sorvolo sintetici e, per gli splat, i formati nativi delle librerie di rendering.

I tempi di elaborazione sono compatibili con scenari operativi: ad esempio sulla scena campione della statua di Piazza Cavour in Ancona (Figura 2) con 121 fotogrammi registrati l'intera pipeline, dall'ingestione del video al modello georiferito pronto per la pubblicazione, si completa in circa 25 minuti su una workstation dotata di una singola GPU commerciale (NVIDIA RTX 5070). La stessa pipeline inoltre è installata in versione prototipale sul cluster di calcolo Margherita di Regione Emilia-Romagna per l'elaborazione massiva e concorrente di più flussi video.

3. Georeferenziazione metrica automatica

I modelli prodotti dalla Structure-from-Motion sono definiti a meno di una similitudine: scala, orientamento e posizione assoluti sono indeterminati. SplatTwin li stima automaticamente dalla telemetria del volo, senza punti di appoggio a terra. La traccia GPS è estratta dai formati più diffusi quali sottotitoli SRT dei droni DJI, tracce GPX e NMEA, flight record binari delle app di volo e sincronizzata temporalmente con i fotogrammi; per le sorgenti prive di sincronizzazione nativa l'offset temporale è stimato per ricerca esaustiva, minimizzando il residuo di allineamento. Un allineamento di similarità robusto (stima di Umeyama [6] con reiezione iterativa degli outlier) fra la traiettoria delle camere e la traccia GPS proiettata in un riferimento locale ENU restituisce la trasformazione completa del modello in coordinate metriche.

Per la componente altimetrica la quota GNSS è riportata dall'altezza sul livello medio del mare all'altezza ellissoidi ca WGS84 mediante il modello di geoide EGM96; quando il dato altimetrico di bordo risulta inaffidabile, eventualità riscontrata su alcuni firmware, la quota del modello viene agganciata automaticamente al modello digitale del terreno. La nuvola densa viene infine ripulita (rimozione degli artefatti di cielo e degli outlier statistici, tipicamente il 25–30% dei punti esportati) ed esportata in formato .las georiferito, corredata di un file di metadati con tutte le trasformazioni applicate.

Sui voli di test l'allineamento converge con residui mediani di 0,15 m su traiettorie di 60–100 m, pari allo 0,2–0,3% dell'estensione del rilievo, con il 100% dei fotogrammi classificati inlier: l'errore di scala del modello risulta quindi dell'ordine del 1%. La precisione relativa delle misure effettuate sul modello (distanze, aree) è tipicamente del 1–2%, verificata confrontando elementi noti della scena con le ortofoto; l'accuratezza assoluta di posizionamento resta invece limitata dal ricevitore GNSS di bordo (1–3 m in planimetria per i droni consumer), migliorabile in prospettiva con punti di appoggio o ricevitori RTK. Sono in corso studi più approfonditi per misurare con maggior precisione gli errori di misura e posizionamento assoluti.

4. Pubblicazione: 3D Tiles e Cesium ion

Il modello georiferito viene anche pubblicato automaticamente come 3D Tiles [7], lo standard OGC per lo streaming di contenuti 3D geospaziali. In questa fase la piattaforma cloud Cesium ion svolge il ruolo di canale di stoccaggio e distribuzione: la pipeline carica la nuvola .las (o la mesh) con la posizione geografica calcolata, il servizio esegue il tiling e il modello risulta immediatamente consultabile e misurabile via browser, correttamente posizionato su terreno e ortofoto, affiancabile agli altri strati informativi di un qualunque client CesiumJS. Lo stesso flusso è integrabile con le infrastrutture di dati territoriali esistenti, a partire dal Geoportale della Regione Emilia-Romagna [9], che già espone i propri strati secondo standard OGC.

Cesium ion non è però soltanto un canale di distribuzione: la piattaforma offre a sua volta un servizio di ricostruzione e tiling 3D Gaussian Splatting a partire dalle immagini. È quindi naturale confrontare il risultato del servizio cloud con quello della nostra implementazione locale. Le Figure 2 e 3 mostrano il confronto sul rilievo della statua di Piazza Cavour in Ancona: la qualità visiva dei due risultati è molto simile; la differenza sostanziale è nei tempi e nel controllo del processo, con SplatTwin che completa la ricostruzione in circa 25 minuti su GPU commerciale mentre il servizio cloud richiede tempi sensibilmente maggiori, dipendenti dalla coda del servizio e non controllabili dall'utente. Sono in corso studi più approfonditi per misurare in dettaglio le differenze di risultato visuale mediante metriche oggettive.

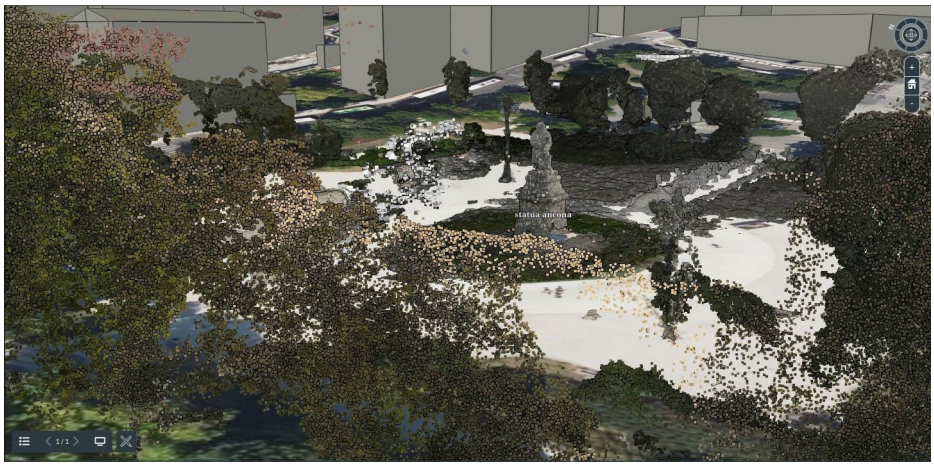

Fig. 2. Nuvola di punti georiferita della statua di Piazza Cavour (Ancona) pubblicata come 3D Tiles su Cesium ion e visualizzata sul globo virtuale, correttamente posizionata e in scala rispetto a ortofoto, terreno ed edifici OSM.

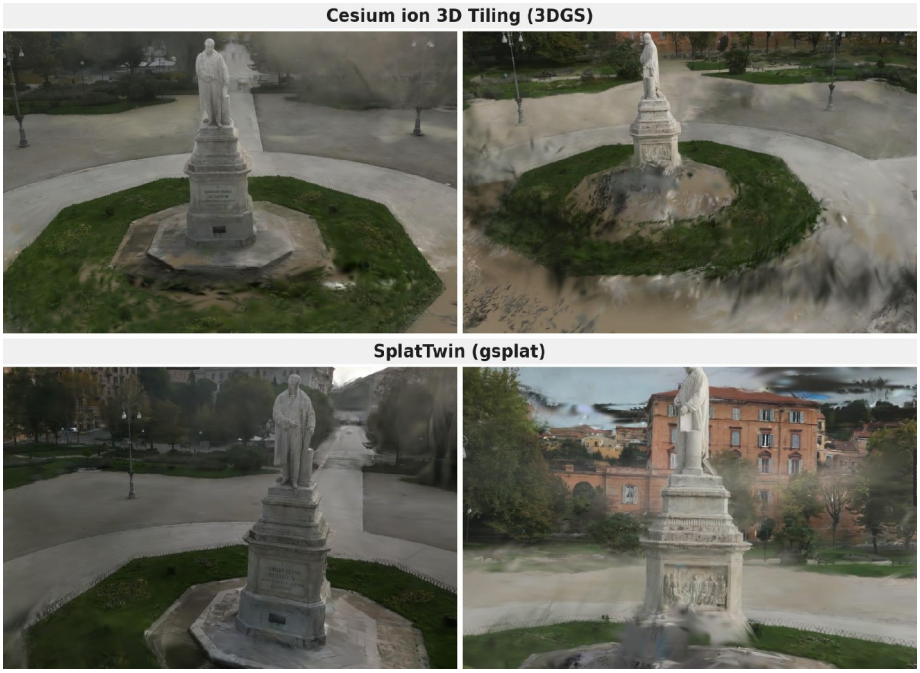

Fig. 3. Confronto qualitativo sul rilievo della statua: tiling 3D Gaussian Splatting di Cesium ion in qualità alta (in alto) e rendering dell'implementazione SplatTwin basata sulla libreria gsplat (in basso), vista frontale e laterale. I risultati visuali sono molto simili a livello qualitativo.

Un secondo elemento di valutazione è la titolarità del dato. I termini di servizio di Cesium ion (sez. 2.6, License & Warranty to Cesium) [10] prevedono che l'utente conceda a Cesium una licenza non esclusiva, mondiale, trasferibile e sub-licenziabile per usare, copiare, modificare, trasformare, distribuire e memorizzare i dati caricati e i prodotti derivati, anche al fine di migliorare i servizi della piattaforma. Con SplatTwin l'intera elaborazione dal video sorgente al modello georiferito avviene all'interno dell'infrastruttura dell'ente (workstation locale o cluster Margherita): il proprietario mantiene la sovranità sul proprio dato che non abbandona mai il perimetro di proprietà. Infatti la pubblicazione su piattaforme esterne è un passo finale opzionale, limitato ai soli prodotti derivati che si desidera effettivamente distribuire.

5. Casi di studio

L'approccio è stato validato su scenari volutamente eterogenei per scala e contenuto. Oltre al già citato caso della statua di Piazza Cavour ad Ancona (Figure 2 e 3) con 5 milioni di punti esportati, ridotti a 3,5 milioni dopo la pulizia automatica.

La Figura 4 mostra appunto il rendering di una frana nell'Appennino Romagnolo, generato dal modello addestrato su un volo drone.



Fig. 4. Esempio di rendering grafico 3D ad alta risoluzione rappresentante la porzione di una frana nell'Appennino Romagnolo: immagine generata tramite genAI a partire da un volo drone.

Alla scala del singolo monumento, la Figura 5 mostra invece la nuvola di punti della scultura “Bimba sull'altalena” nel centro storico di Cesena (FC): strutture sottili come le corde dell'altalena e la vegetazione circostante vengono ricostruite correttamente, confermando l'idoneità del metodo alla documentazione del patrimonio urbano. Alla scala territoriale, la Figura 6 mostra la ricostruzione di un rilievo montano a partire da un volo relativo all'estate 2026: la morfologia del crinale e le variazioni di copertura del versante risultano leggibili nonostante l'assenza di strutture antropiche di riferimento.

La Figura 7 torna sul tema emergenziale con la nuvola di punti dell'abitato di Faenza (RA) colpito dall'alluvione, ricostruita dalle riprese effettuate durante l'evento: edifici, corso d'acqua e aree esondate sono distinguibili e misurabili, a conferma che anche riprese non pianificate e in condizioni operative difficili costituiscono una sorgente dati sufficiente per una prima valutazione quantitativa del danno.



Fig. 5. Nuvola di punti della scultura Bimba sull'altalena (Cesena): le strutture sottili dell'altalena e il contesto murario sono ricostruiti correttamente da un breve video ravvicinato.

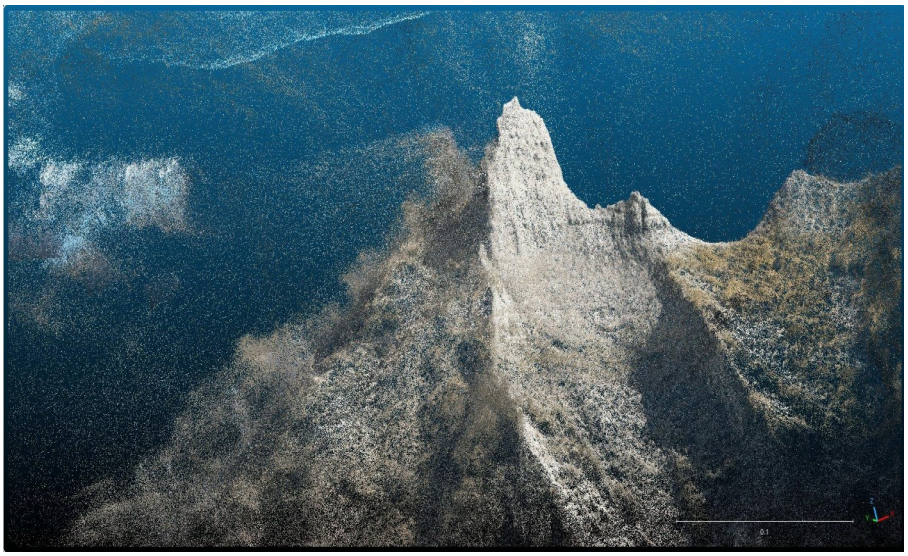

Fig. 6. Ricostruzione di un rilievo montano da volo drone: la morfologia del crinale è leggibile anche in assenza di riferimenti antropici.

Infine, un caso dedicato alle infrastrutture: il portale di una galleria ferroviaria nei pressi di Brisighella (RA), ostruito da un movimento franoso. Dal video di ispezione sono stati estratti 92 fotogrammi full HD (1920×1080); l'addestramento NeRF, 20.000 iterazioni, si completa in 14 minuti su una RTX 5070 e in 18 minuti su una RTX 3070 Super: l'intera pipeline resta quindi praticabile anche su GPU commerciali di generazione precedente. La Figura 8 affianca un fotogramma del rendering NeRF alla

nuvola di punti georiferita: portale, corpo di frana e versante sono documentati con dettaglio sufficiente per una prima valutazione dello stato dell'opera, senza accesso diretto all'area.



Fig. 7. Nuvola di punti dell'abitato di Faenza colpito dall'alluvione, ricostruita dalle riprese effettuate durante l'emergenza: aree esondate ed edifici sono distinguibili e misurabili.



Fig. 8. Galleria ferroviaria presso Brisighella (RA): rendering NeRF e nuvola di punti georiferita, da 92 fotogrammi full HD; addestramento NeRF di 20.000 iterazioni in 18 minuti (RTX 3070 Super), 14 minuti (RTX 5070) e 7,5 minuti (RTX4090).

6. Utilità per l'utilizzatore

Dal punto di vista dell'utilizzatore finale, per il tecnico di un ente territoriale, di una protezione civile o di uno studio professionale i vantaggi operativi sono concreti: (i) non servono sensori dedicati né missioni di rilievo pianificate: è sufficiente un video di drone, anche amatoriale o di repertorio; (ii) il modello è pronto in decine di minuti su hardware commerciale, non in giorni; (iii) la fruizione avviene nel browser, senza software specialistico: il modello è navigabile e misurabile direttamente sul globo virtuale, sovrapposto a terreno, ortofoto ed edificato; (iv) il dato resta sotto il controllo dell'ente, requisito spesso vincolante per le pubbliche amministrazioni; (v) i modelli georiferiti sono riutilizzabili nei flussi GIS esistenti tramite i formati standard .las e 3D Tiles.

7. Conclusioni

Abbiamo presentato SplatTwin, una pipeline completa che trasforma un video di volo drone in un gemello digitale georiferito, pubblicato automaticamente come 3D Tiles e consultabile via browser. La combinazione di 3D Gaussian Splatting su libreria gsplat, georeferenziazione automatica dalla telemetria e pubblicazione integrata porta i tempi complessivi a circa 25 minuti su GPU commerciale, con errori di scala dell'ordine del 1% e precisione relativa delle misure del 1–2%. I casi di studio di monumenti urbani, versanti montani e aree alluvionate confermano l'applicabilità del metodo a contesti molto diversi, con particolare valore nei contesti dinamici ed emergenziali. Gli sviluppi futuri riguardano il miglioramento dell'accuratezza assoluta tramite punti di appoggio e ricevitori RTK, lo streaming diretto degli splat con livelli di dettaglio, e l'integrazione sistematica dei modelli prodotti nelle infrastrutture di dati territoriali regionali. Sono in corso studi mirati per misurare con precisione la resa visuale, geometrica e di georeferenziazione del metodo proposto.

Riferimenti bibliografici

1. Pesic, M., Lughi, G., and Roffilli, M.: Intelligenza Artificiale Generativa per Digital Twin Urbanistico partecipato, ASITA 2024
2. Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R.: NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. ECCV 2020, arXiv:2003.08934.
3. Mazzacca, G., Karami, A., Rigon, S., Farella, E. M., Trybala, P., and Remondino, F.: NERF for heritage 3D reconstruction. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLVIII-M-2-2023, 1051-1058 (2023).

4. Huang, B., Yu, Z., Chen, A., Geiger, A., and Gao, S.: 2D Gaussian Splatting for Geometrically Accurate Radiance Fields. arXiv:2403.17888 (2024).
5. Kerbl, B., Kopanas, G., Leimkühler, T., and Drettakis, G.: 3D Gaussian Splatting for Real-Time Radiance Field Rendering. ACM Trans. Graph. 42(4), 139 (2023).
6. Umeyama, S.: Least-Squares Estimation of Transformation Parameters Between Two Point Patterns. IEEE Trans. Pattern Anal. Mach. Intell. 13(4), 376–380 (1991).
7. Open Geospatial Consortium: 3D Tiles Specification 1.1, OGC 22-025r4 (2023).
8. Ye, V., Li, R., Kerr, J., Turkulainen, M., Yi, B., Pan, Z., Seiskari, O., Ye, J., Hu, J., Tancik, M., and Kanazawa, A.: gsplat: An Open-Source Library for Gaussian Splatting. Journal of Machine Learning Research 26(34), 1–17 (2025).
9. Regione Emilia-Romagna: Geoportale l'infrastruttura regionale dei dati territoriali. https://geoportale.regione.emilia-romagna.it (consultato ottobre 2026).