

Interpretazione della scena e localizzazione in ambito urbano per veicoli autonomi

Augusto Luis Ballardini, Danilo Benetti, Daniele Cattaneo, Simone Fontana, Matteo Vaghi and
Domenico Giorgio Sorrenti

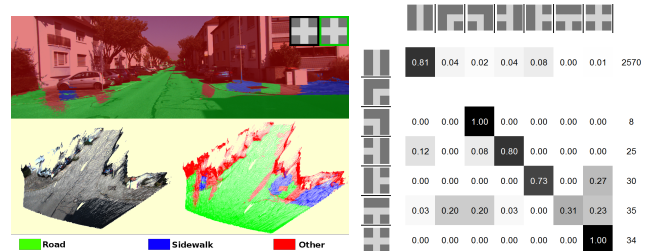
Università degli Studi di Milano - Bicocca, Milano, Italy
{ballardini, daniele.cattaneo, simone.fontana, sorrenti}@disco.unimib.it
{d.benetti1,m.vaghi9}@campus.unimib.it

Abstract

Nel campo della robotica mobile la localizzazione dell'osservatore nel suo ambiente operativo costituisce da sempre un problema chiave. La ricerca condotta nel lab. IRA (Informatica e Robotica per l'Automazione) del Dipartimento di Informatica, Sistemistica e Comunicazione dell'Università degli Studi di Milano - Bicocca si colloca nel settore dell'intelligenza artificiale per *Sistemi di Trasporto* e *"Smart Cities"*.

1 Introduzione

Nel corso degli ultimi anni, il lab. IRA si è dedicato ai sistemi di localizzazione in ambito urbano per veicoli a guida autonoma, basandosi sulla identificazione e riconoscimento di componenti stradali sia di basso livello, come ad esempio carreggiata o segnaletica orizzontale, sia di alto livello semantico come ad esempio posizione e geometria di un incrocio stradale o le facciate degli edifici. Se, per quanto concerne il trasporto autostradale, le applicazioni del settore automotive hanno raggiunto livelli di affidabilità tali da rendere già disponibili sul mercato sistemi di guida autonoma a livello SAE-2, le condizioni sono decisamente diverse nei centri urbani. In questi casi la corretta stima della posizione e orientamento del veicolo rispetto all'ambiente circostante è un problema ancora aperto. Ciò è dovuto principalmente alla inaffidabilità ed alla scarsa robustezza dei sistemi di posizionamento globale quali GALILEO o GPS. Nel lab. IRA abbiamo un particolare interesse per l'integrazione di mappe stradali pre-esistenti, inizialmente concepite per l'utilizzo da parte dell'uomo, quali quelle fornite da servizi come Google, Bing, Here e progetti open-source quali OpenStreetMap. L'idea consiste nello sviluppo di algoritmi di intelligenza artificiale capaci di sfruttare l'enorme quantità di informazione contenuta in questi servizi online, mediante una ricerca di corrispondenze con i dati sensoriali ottenuti a bordo veicolo. Le applicazioni di riferimento sono quelle del trasporto pubblico, sia di persone che di merci. Si prevede per prossimo futuro un cambio del paradigma di mobilità urbana mediante la graduale introduzione di flotte di micro-bus autonomi chiamabili a richiesta dall'utente, ad esempio via smartphone, secondo uno schema assimilabile ad un servizio Taxi, su scala molto maggiore. Questa idea trova applicazione sia nel trasporto delle persone che nella consegna di merci in area urbana, ad esempio per il collegamento tra i grandi interporti ed i singoli esercizi commerciali citta-



(a) Esempio di esecuzione del rilevatore di incroci su un frame del dataset KITTI, in cui la topologia viene classificata correttamente. (b) Matrice di confusione derivata da un dataset in cui l'approccio proposto ha ottenuto risultati promettenti.

Figura 1

dini. In quest'ottica i principali progetti possono essere così riassunti.

- Creazione del framework probabilistico per l'interpretazione della scena.
- Riconoscimento della geometria e localizzazione del veicolo utilizzando incroci stradali.
- Stima della posizione del veicolo mediante il riconoscimento delle facciate degli edifici.
- Localizzazione mediante associazione di immagini a mappe prodotte con sensori LiDAR.

2 Framework per l'interpretazione della scena

Lo scopo del lavoro rientra nel progetto denominato Road Layout Estimation (RLE). Esso consiste nello studio e sviluppo di un sistema di comprensione della scena osservata da una macchina a guida autonoma, ovvero un sistema probabilistico per la stima della scena circostante mediante una analisi degli elementi caratterizzanti. Il lavoro costituisce l'inizio di un ambizioso progetto, volto a racchiudere in un'unica piattaforma i principali aspetti della robotica e computer vision, al fine di permettere la navigazione autonoma di un veicolo in ambiente urbano superando alle imprecisioni dei sistemi satellitari di navigazione globale (GNSS). Attualmente il sistema è in grado di sfruttare il riconoscimento di diversi elementi, sia fisici, come gli edifici, che concettuali, come la geometria di un incrocio. Gli elementi sono confrontabili con la cartografia fornita dal sistema OpenStreetMap, che può essere utilizzata come base di dati per la localizzazione del veicolo. Il lavoro è stato presentato in [Ballardini *et al.*, 2015].

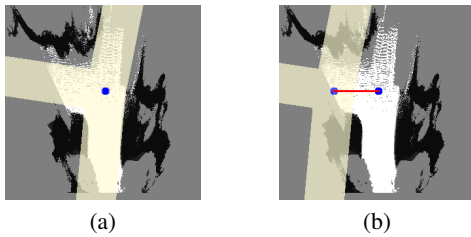


Figura 2: Nella figura, due ipotesi di localizzazione del veicolo rispetto al centro dell'incrocio indicato dal punto blu. In beige, la geometria e posizione dell'incrocio indicati nella cartografia OpenStreetMap. In bianco e nero, l'incrocio identificato dall'algoritmo proposto. Un migliore allineamento tra i due corrisponde ad una migliore stima di posizione del veicolo.

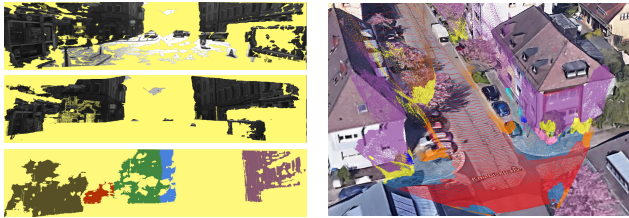


Figura 3: A destra, le varie fasi necessarie ad identificare le facciate degli edifici con l'approccio [Ballardini *et al.*, 2016]. A sinistra, il risultato della pipeline di segmentazione e ricostruzione 3D aggiornata con tecniche basate su reti neurali.

3 Riconoscimento della geometria degli incroci

Nata inizialmente come componente per il progetto RLE, questa ricerca ha portato allo sviluppo di un rilevatore probabilistico di intersezioni stradali in grado di ricostruire la geometria di un incrocio urbano mediante le informazioni provenienti da una coppia di camere montate su un veicolo. Rappresenta il nostro primo approccio verso l'analisi semantica della scena, mediante l'utilizzo combinato di classificazioni geometriche e pixel-level. L'obiettivo è di superare le limitazioni delle tecniche presenti in letteratura, sviluppando una tecnica per il riconoscimento della geometria degli incroci stradali eseguibile *online* ovvero direttamente sul veicolo durante l'approccio all'incrocio. Questa caratteristica si contrappone ai lavori di letteratura della fine degli anni '80. L'approccio, inizialmente presentato in [Ballardini *et al.*, 2017], è stato ulteriormente raffinato con l'utilizzo di reti neurali convoluzionali, come illustrato in Figura 2, e verrà presentato alla conferenza ICRA 2019.

4 Utilizzo delle facciate di edifici come elementi cartografici

Lo scopo di questo modulo è creare un sistema in grado di riconoscere gli edifici in un contesto urbano ed usare questa informazione per contribuire alla localizzazione. Nuovamente, questo ha richiesto la definizione di un algoritmo di computer vision in grado di segmentare gli edifici presenti nella scena, seguito dalla definizione di una procedura per la stima della verosimiglianza di un'ipotesi di localizzazione rispetto alla scena percepita dal veicolo. Per quanto riguarda il riconoscimento delle facciate degli edifici, in letteratura sono



Figura 4: Tre immagini di profondità ottenute tramite raytracing applicato ad una mappa a voxel (octomap): a sinistra l'immagine perturbata secondo un errore iniziale sulla pose, al centro l'immagine ottenuta andando a correggere la pose precedente mediante la predizione della rete neurale, a destra la groundtruth.

presenti sia approcci basati sull'analisi di immagini singole sia approcci che si avvalgono della stereo visione. L'originale algoritmo basato sull'algoritmo di ricostruzione SGBM [Ballardini *et al.*, 2016] è stato ad oggi migliorato con l'utilizzo di reti neurali sia per quanto concerne la ricostruzione tridimensionale che per la segmentazione, vedi Figura 3.

5 Localizzazione mediante associazione di immagini e mappe LiDAR

Negli ultimi anni, nel campo della computer vision, hanno mostrato una grande superiorità gli approcci basati su reti neurali convoluzionali che affrontano alcune problematiche della robotica mobile come, ad esempio, il riconoscimento di ostacoli, la segmentazione semantica delle immagini, la classificazione di cartelli stradali, la ricostruzione tridimensionale e la localizzazione, sfruttando sia i dati catturati dalle camere che dati provenienti da sensori laser. Per quanto concerne il problema della localizzazione, i principali approcci presenti in letteratura sfruttano reti neurali progettate con lo scopo di memorizzare ed integrare al loro interno una area geografica. A questo proposito, i nostri contributi si contraddistinguono grazie all'utilizzo congiunto di immagini e mappe basate su scansioni LiDAR pre-esistenti. L'elemento caratterizzante è quindi la generalizzazione rispetto alla mappa, allo scopo di rendere la rete capace di localizzarsi in luoghi differenti (ovvero su mappe differenti) rispetto a quelle mostrate in fase di training, vedi Figura 4. I nostri approcci consentiranno di sfruttare per intero le potenzialità offerte dalla imminente introduzione di mappe cartografiche ad alta definizione.

Riferimenti bibliografici

- [Ballardini *et al.*, 2015] A. L. Ballardini, S. Fontana, A. Furlan, D. Limongi, e D. G. Sorrenti. A framework for outdoor urban environment estimation. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems (ITSC)*, Sept 2015.
- [Ballardini *et al.*, 2016] A. L. Ballardini, D. Cattaneo, S. Fontana, e D. G. Sorrenti. Leveraging the osm building data to enhance the localization of an urban vehicle. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, Nov 2016.
- [Ballardini *et al.*, 2017] A. L. Ballardini, D. Cattaneo, S. Fontana, e D. G. Sorrenti. An online probabilistic road intersection detector. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, May 2017.