Sistemi di Visione Artificiale Indossabili

Giovanni Maria Farinella, Antonino Furnari
Dipartimento di Matematica e Informatica, Università di Catania
FVP@IPLAB: http://iplab.dmi.unict.it/fpv/
Contatto: gfarinella@dmi.unict.it

Abstract

Dispositivi indossabili quali Microsoft HoloLens e Google Glass Enterprise Edition hanno un grande potenziale d'utilizzo in contesti industriali. Sebbene il mercato già offra dispositivi indossabili adatti all'uso in tali contesti, lo sviluppo di tecniche di intelligenza artificiale nell'ambito applicativo considerato resta un campo relativamente poco esplorato. Questo documento presenta le ricerche condotte dall'Image Processing Laboratory (IPLAB) dell'Università di Catania sulla costruzione di algoritmi di Machine Learning e Computer Vision per dispositivi indossabili, con particolare riferimento al loro utilizzo in contesti di natura industriale.

1 Introduzione

L'utilizzo dell'intelligenza artificiale in contesti industriali promette di incrementare la produttività e migliorare la sicurezza nei luoghi di lavoro. In particolare, dispositivi indossabili capaci di acquisire e analizzare informazioni sull'ambiente circostante, quali immagini e video, hanno un grande potenziale per la costruzione di sistemi intelligenti capaci di accompagnare il lavoratore e assisterlo durante le fasi di lavoro. Nonostante il mercato oggi offra dispositivi quali Microsoft HoloLens e Google Glass Enterprise Edition, che si prestano già all'utilizzo in ambienti industriali, l'applicazione di algoritmi di intelligenza artificiale nel contesto dei dispositivi indossabili dotati di visione è un campo poco esplorato.

Questo documento presenta le ricerche condotte dall'Image Processing Laboratory (IPLAB) dell'Università degli Studi di Catania nell'ambito della costruzione di algoritmi di Machine Learning e Computer Vision per dispositivi indossabili, con particolare riferimento all'utilizzo di tali tecnologie in contesti industriali. In particolare, verranno presentate le ricerche condotte in tre ambiti di rilevanza industriale, ovvero la *localizzazione* basata su immagini acquisite da dispositivi indossabili, il *riconoscimento di oggetti e azioni* e l'*anticipazione* (ovvero la predizione in anticipo) di interazioni con oggetti e di azioni. Il lettore è rimandato alla pagina http://iplab.dmi.unict.it/fpv/ per maggiori informazioni sulle ricerche condotte dal laboratorio IPLAB nel contesto della visione artificiale in "prima persona" (First Person Vision).

2 Localizzazione

La capacità di determinare la posizione da immagini acquisite da dispositivi indossabili permette di localizzare gli operatori all'interno di un edificio, di fornire informazioni contestualizzate e guidarli verso una destinazione e di adattare il comportamento degli algoritmi sulla base del contesto.

L'IPLAB si è occupato del riconoscimento dell'ambiente in cui l'utente opera a livello di "personal location" [Furnari *et al.*, 2017c; Furnari *et al.*, 2018b], ovvero ambienti delimitati all'interno dei quali l'utente svolge determinate attività. Esempi di "personal locations" sono l'ufficio e il banco da lavoro. L'insieme di "personal location" rilevanti dipende dall'utente, pertanto gli algoritmi sviluppati permettono di riconoscere gli ambienti a partire da un insieme limitato di campioni di training forniti dall'utente stesso. Dato l'utente potrebbe trovarsi in un ambiente che non è stato specificato in fase di setup, gli algoritmi sviluppati sono in grado di riconoscere gli ambienti di interesse scartando gli altri ambienti (i "negativi"). Oltre a permettere la localizzazione in tempo reale, gli algoritmi sviluppati consentono di segmentare un video in unità temporali coerenti basate sul contesto in cui l'utente opera. Il sistema di localizzazione è stato anche applicato nel contesto del riconoscimento di ambienti per la localizzazione dei visitatori di un museo [Ragusa *et al.*, 2019b].[1]

IPLAB ha inoltre considerato il problema della localizzazione al livello della stima della posa della camera. Nello specifico, ci si è occupati della localizzazione di carrelli della spesa in un supermercato da immagini acquisite con telecamere montate sui carrelli [Spera *et al.*, 2018]. Localizzare i carrelli permette di costruire sistemi intelligenti capaci di guidare i clienti nel punto vendita e studiare il loro comportamento per fornire servizi personalizzati [Santarcangelo *et al.*, 2018]. La localizzazione viene effettuata con tecniche di image retrieval, avvalendosi di una metrica costruita mediante deep metric learning.[2] Le stesse tecnologie possono essere utilizzate per la costruzione di sistemi capaci di localizzare gli operatori e guidarli all'interno di un magazzino o di altro ambiente industriale. La stima della posa della camera da dispositivi indossabili è stata investigata anche nel contesto di dati simulati a partire dal modello 3D di un edificio reale [Orlando *et al.*, 2019]. Generare dati sintetici permette di

[1] Video dimostrativo: https://youtu.be/VYZ6Awqy1ko
[2] Video dimostrativo: https://youtu.be/BxbdgWxFhgc

ottenere grandi quantità di dati etichettati, adatti per lo studio di algoritmi di localizzazione.

3 Riconoscimento di Oggetti e Azioni

Riconoscere gli oggetti con i quali l'utente interagisce e le azioni effettuate da immagini acquisite mediante dispositivi indossabili è utile per comprendere gli obiettivi dell'utente e verso cosa egli ripone l'attenzione. In questo contesto, sono stati investigati algoritmi per individuare i punti di interesse in un museo [Ragusa *et al.*, 2019a]. Gli algoritmi sviluppati permettono di riconoscere gli oggetti presenti nella scena e stimare quali tra essi sono attualmente osservati dall'utente.[3]

E' stato inoltre studiato il problema della segmentazione temporale di video sulla base delle azioni effettuate dagli utenti [Furnari *et al.*, 2017a]. Gli algoritmi investigati permettono di identificare i tempi di inizio e fine di ogni azione in un video, scartando i frame "negativi" in corrispondenza dei quali nessuna azione viene compiuta dall'utente.

Per favorire lo studio di problemi di riconoscimento e predizione anticipata di oggetti e azioni, l'IPLAB ha collaborato alla creazione di EPIC-KITCHENS [Damen *et al.*, 2018], uno dei più grandi dataset di video acquisiti in prima persona per il riconoscimento di azioni, l'anticipazione di azioni e il riconoscimento di oggetti. Il dataset è stato acquisito da 32 soggetti in 3 diverse nazioni (Italia, Regno Unito e Canada) e contiene 55 ore di video, annotazioni per circa 40000 azioni e 500000 oggetti.

4 Anticipazione di Interazioni con Oggetti

Una caratteristica desiderabile per un dispositivo indossabile dotato di intelligenza artificiale, è la capacità di prevedere in anticipo cosa accadrà nella scena. Tale capacità permette di costruire sistemi che possono guidare l'utente in flussi di lavoro complessi e notificarlo nel caso in cui una azione sbagliata o pericolosa sta per essere intrapresa.

L'IPLAB ha investigato algoritmi per predire quali tra gli oggetti presenti nella scena verrà utilizzato dall'utente nel breve termine. In particolare, gli studi condotti [Furnari *et al.*, 2017b] hanno evidenziato come l'analisi delle traiettorie di oggetti individuati da video acquisiti in prima persona, permetta di ottenere informazioni sui prossimi oggetti che verranno utilizzati dall'utente in un contesto dinamico.[4]

Il tema dell'anticipazione di interazioni con oggetti è stato anche investigato con la definizione di una challenge su "egocentric action anticipation" in relazione al dataset EPIC-KITCHENS [Damen *et al.*, 2018] e con lo studio di architetture di AI e misure di valutazione idonee ad affrontare il problema della predizione anticipata di azioni da video in prima persona [Furnari *et al.*, 2018a]. Gli algoritmi sviluppati permettono di predire l'insieme delle prossime azioni probabili a partire dall'osservazione di video prima che esse accadano.[5]

[3] Video dimostrativo: https://youtu.be/nBkYOdKYu0s

[4] Video: http://iplab.dmi.unict.it/NextActiveObjectPrediction/

[5] Video dimostrativo: https://youtu.be/w_3FiIcnUlc

5 Conclusione

Questo documento ha presentato le ricerche condotte da IPLAB nell'ambito della costruzione di algoritmi di intelligenza artificiale per dispositivi di visione indossabili. I problemi affrontati hanno potenziale applicativo in contesti industriali e gravitano attorno a tre temi principali relativi alla *localizzazione basata su immagini*, al *riconoscimento di oggetti e azioni* e all'*anticipazione di interazioni con oggetti*.

Riferimenti bibliografici

[Damen *et al.*, 2018] D. Damen, H. Doughty, G. M. Farinella, S. Fidler, A. Furnari, E. Kazakos, D. Moltisanti, J. Munro, T. Perrett, W. Price, e M. Wray. Scaling egocentric vision: The epic-kitchens dataset. In *European Conference on Computer Vision*, 2018.

[Furnari *et al.*, 2017a] A. Furnari, S. Battiato, e G. M. Farinella. How shall we evaluate egocentric action recognition? In *ICCV Workshops*, 2017.

[Furnari *et al.*, 2017b] A. Furnari, S. Battiato, K. Grauman, e G. M. Farinella. Next-active-object prediction from egocentric videos. *Journal of Visual Communication and Image Representation*, 49:401 – 411, 2017.

[Furnari *et al.*, 2017c] A. Furnari, G. M. Farinella, e S. Battiato. Recognizing personal locations from egocentric videos. *IEEE Transactions on Human-Machine Systems*, 47(1):6–18, 2017.

[Furnari *et al.*, 2018a] A. Furnari, S. Battiato, e G. M. Farinella. Leveraging uncertainty to rethink loss functions and evaluation measures for egocentric action anticipation. In *ECCV Workshop on Egocentric Perception, Interaction and Computing (EPIC)*, 2018.

[Furnari *et al.*, 2018b] A. Furnari, S. Battiato, e G. M. Farinella. Personal-location-based temporal segmentation of egocentric video for lifelogging applications. *Journal of Visual Communication and Image Representation*, 52:1–12, 2018.

[Orlando *et al.*, 2019] S. A. Orlando, A. Furnari, S. Battiato, e G. M. Farinella. Image-based localization with simulated egocentric navigations. In *International Conf. on Computer Vision Theory and Applications*, 2019.

[Ragusa *et al.*, 2019a] F. Ragusa, A. Furnari, S. Battiato, G. Signorello, e G. M. Farinella. Egocentric point of interest recognition in cultural sites. In *International Conf. on Computer Vision Theory and Applications*, 2019.

[Ragusa *et al.*, 2019b] F. Ragusa, A. Furnari, S. Battiato, G. Signorello, e G. M. Farinella. Egocentric visitors localization in cultural sites. *Journal on Computing and Cultural Heritage*, 2019.

[Santarcangelo *et al.*, 2018] V. Santarcangelo, G. M. Farinella, A. Furnari, e S. Battiato. Market basket analysis from egocentric videos. *Pattern Recognition Letters*, 112:83–90, 2018.

[Spera *et al.*, 2018] E. Spera, A. Furnari, S. Battiato, e G. M. Farinella. Egocentric shopping cart localization. In *International Conference on Pattern Recognition*, 2018.