

ENIGMA: Egocentric Navigator for Industrial Guidance, Monitoring and Anticipation

Francesco Ragusa^{1,2,*}, Antonino Furnari^{1,2}, Antonino Lopes³, Marco Moltisanti³, Emanuele Ragusa³, Marina Samarotto³, Luciano Santo³, Nicola Picone⁴, Leo Scarso⁴ and Giovanni Maria Farinella^{1,2}

¹FPV@IPLAB, DMI - University of Catania, Italy

²Next Vision s.r.l. - Spinoff of the University of Catania, Italy

³Xenia Gestione Documentale s.r.l. - Xenia Progetti s.r.l., Acicastello, Catania, Italy

⁴Morpheos s.r.l. - Catania, Italy

Abstract

Il sistema ENIGMA (Egocentric Navigator for Industrial Guidance, Monitoring and Anticipation) è un sistema integrato di supporto ai lavoratori che operano in laboratori industriali. Il sistema ENIGMA include un assistente indossabile che interpreta il comportamento dei lavoratori grazie ad algoritmi di Computer Vision che 1) localizzano l'operatore nello spazio 2) riconoscono gli oggetti presenti nel laboratorio 3) individuano le interazioni tra l'operatore e gli oggetti 4) anticipano quale sarà il prossimo oggetto con cui l'operatore interagirà. Inoltre, è dotato di un backend che estrae informazioni semantiche di alto livello sul comportamento dell'operatore al fine di fornire servizi utili al miglioramento della sua sicurezza. Gli esperimenti preliminari mostrano buoni risultati per quanto riguarda la localizzazione, l'individuazione e il riconoscimento di oggetti e il rilevamento di interazioni da camere in prima persona.

Keywords

Egocentric Vision, First Person Vision, Industrial Domain

1. Introduzione

La comprensione del comportamento umano da un punto di ripresa solidale con il soggetto operante (*egocentric*) consente lo sviluppo di sistemi intelligenti integrati (ad es. smartglasses) capaci di supportare gli esseri umani nella loro vita quotidiana in diversi scenari, quali l'ambiente domestico [1], i siti culturali [2, 3] ed edifici industriali [4, 5, 6].

In quest'ultimo scenario, in particolare, la localizzazione degli utenti può risultare estremamente utile nella gestione di emergenze quali incendi e terremoti, fornendo ai soccorritori indicazioni puntuali sull'ultima posizione nota delle persone in ambienti indoor. Anche il riconoscimento degli oggetti risulta utile, in quanto consente di dare informazioni aggiuntive sul loro utilizzo ed eventuali avvertenze, e abilita lo sviluppo di algoritmi che consentono di rilevare le interazioni tra gli oggetti stessi e l'operatore. Questi ultimi portano notevoli vantaggi, ad esempio fornendo indicazioni sulla corretta esecuzione di procedure complesse o consentendo l'implementazione di strategie per il risparmio energetico. Inoltre, anticipare quale sarà il prossimo oggetto con cui l'operatore interagirà consente di aumentare la sicurezza dell'operatore

stesso, ad esempio mostrando all'utente una notifica nel caso l'oggetto con cui si sta per interagire sia pericoloso.

Nonostante siano già presenti molti sistemi che fanno uso di realtà virtuale [7] o aumentata [8], o anche di assistenza remota [9], al fine di formare gli operatori o di guidarli in una procedura complessa [10, 11], l'assenza di comprensione dell'ambiente circostante e del comportamento dell'operatore limita le potenzialità di tali strumenti. Il sistema ENIGMA (Egocentric Navigator for Industrial Guidance, Monitoring and Anticipation) è un assistente indossabile intelligente capace di supportare l'operatore durante task complessi, attraverso suggerimenti sull'esecuzione di operazioni di manutenzione e riparazione, incrementandone la sicurezza attraverso l'anticipazione di interazioni potenzialmente pericolose. Inoltre consente di ridurre il consumo energetico attraverso l'implementazione di opportune strategie, che si basano sull'interazione con l'ambiente circostante. Questi obiettivi vengono raggiunti attraverso algoritmi che consentono di localizzare l'operatore all'interno del laboratorio, di riconoscere gli oggetti nell'ambiente circostante e di rilevare anche le interazioni dell'operatore con gli stessi, e infine di anticipare quale sarà il prossimo oggetto con cui il lavoratore interagirà (Figura 1).

Il sistema ENIGMA è stato testato in un laboratorio che riproduce fedelmente uno scenario industriale. Nel laboratorio sono presenti 23 oggetti differenti, sia fissi che mobili. Inoltre, sono stati installati dispositivi IoT di controllo collegati sia alle prese di corrente che ai

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29–31, 2023, Pisa, Italy

*Corresponding author.

✉ name.surname@organization.it (F. Ragusa)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

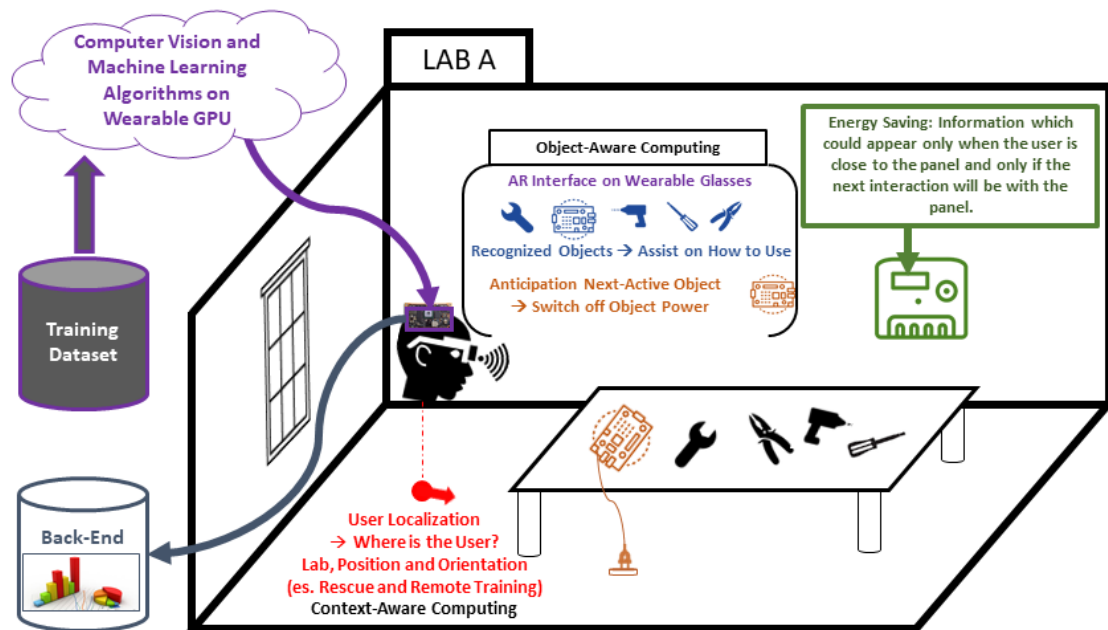


Figure 1: Illustrazione del sistema ENIGMA.

quadri elettrici. Negli esperimenti sono stati considerati 8 contesti, 23 oggetti e 22 differenti interazioni uomo-oggetto. I risultati mostrano che il sistema raggiunge buoni risultati riguardo i task di localizzazione, rilevamento e riconoscimento degli oggetti e interazione tra uomo e oggetto. I futuri esperimenti verteranno sull'anticipazione del prossimo oggetto con cui l'operatore interagirà.

2. Dataset e setting sperimentale

Il laboratorio, mostrato in Figura 2, è stato allestito per riprodurre nella maniera più fedele possibile uno scenario industriale. Sono stati installati dispositivi fissi (quali quadri elettrici, una stazione saldatrice, ecc.) e mobili (ad esempio avvitatori, schede elettriche). Sono stati raccolti ed etichettati due differenti dataset contenenti video in prima persona al fine di acquisire i dati necessari all'implementazione del sistema, e cioè:

1. Localizzazione;
2. Rilevamento e riconoscimento di oggetti;
3. Rilevamento delle interazioni uomo-oggetto;
4. Rilevamento del prossimo oggetto attivo.

2.1. Dataset per la localizzazione

Sono stati acquisiti 62 video utilizzando due dispositivi Microsoft Hololens, con una risoluzione di 2272x1278

pixel ad un frame rate di 30 frame al secondo. Da questi video sono stati estratti 55624 frame, che sono stati successivamente divisi in Training, Validation e Test set, composti rispettivamente da 39437, 4394 e 11993 frame. I frame sono stati etichettati sfruttando COLMAP¹, un software open source che ricostruisce la struttura 3D degli ambienti con un approccio Structure-From-Motion (SfM). In tal modo, si ottiene per ogni frame la corrispondenza tra un punto nello spazio 3D e la sua proiezione sul piano 2D dell'immagine. Questi dati sono stati utilizzati per l'addestramento degli algoritmi e la valutazione delle performance, sia per la localizzazione puntuale (attraverso coordinate cartesiane e orientamento) che contestuale (classificazione del contesto). A tal riguardo, il laboratorio è stato suddiviso in 8 zone, ognuna associata ad un contesto, e specificatamente:

- Ingresso;
- Pannello A;
- Pannello B;
- Estintore;
- Banco da lavoro;
- Armadio;
- Laboratorio, cioè tutto quello che non è incluso nei 7 contesti precedenti.

In Figura 3 è riportata una mappa dei contesti.

¹<https://github.com/colmap/colmap>

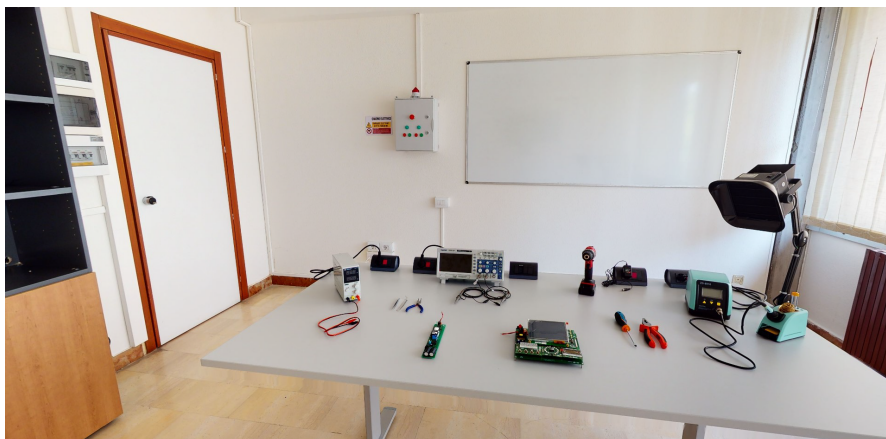


Figure 2: Il laboratorio di test del sistema ENIGMA.

2.2. Dataset delle procedure

Questo dataset è composto da 8 video *egocentric* acquisiti con un dispositivo Microsoft Hololens 2, in cui 7 differenti soggetti effettuano operazioni di riparazione e manutenzione di schede elettroniche nel laboratorio. I video hanno una risoluzione di 2272x1278 pixel, e sono stati acquisiti ad un frame rate di 30 fps. Le annotazioni riguardano le interazioni uomo-oggetto e sono state effettuate manualmente, selezionando il frame in cui l'operatore entra in contatto con l'oggetto e il frame successivo alla terminazione del contatto. La sequenza è stata etichettata con un verbo che descrive meglio l'interazione tra i seguenti quattro:

1. Take (Prendi);
2. Release (Rilascia);
3. Contact (Contatto iniziato);

4. De-contact (Contatto terminato).

Inoltre sono stati etichettati sia l'oggetto coinvolto nell'interazione (oggetto attivo) che tutti gli altri oggetti presenti nella scena. L'etichetta che descrive il frame, quindi, è rappresentata da una tupla con il formato descritto di seguito

$$(x, y, w, h, c, a) \quad (1)$$

in cui:

1. (x, y) sono le coordinate dell'angolo in alto a sinistra del bounding box;
2. (w, h) sono larghezza ed altezza del bounding box;
3. c è la classe dell'oggetto;
4. a è un flag booleano che indica se l'oggetto è attivo nell'interazione.

In Figura 4 sono mostrati alcuni frame di esempio, tra i 20000 etichettati.



Figure 3: Contesti nel laboratorio.

3. Architettura

Il sistema ENIGMA (Figura 5) è composto da 4 parti principali:

- **Dispositivi indossabili:** l'operatore è dotato di dispositivi indossabili, come ad esempio Microsoft Hololens 2. Questi dispositivi consentono al lavoratore di visualizzare le attività di riparazione e collaudo in Realtà Aumentata, avendo allo stesso tempo le mani libere. Inoltre, i video e le immagini acquisite dal dispositivo vengono inviate al modulo di Intelligenza Artificiale per essere elaborate.

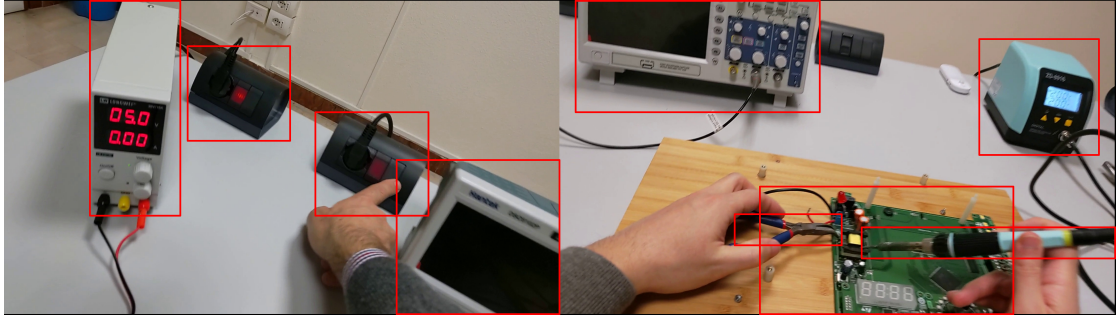


Figure 4: Frame etichettati.

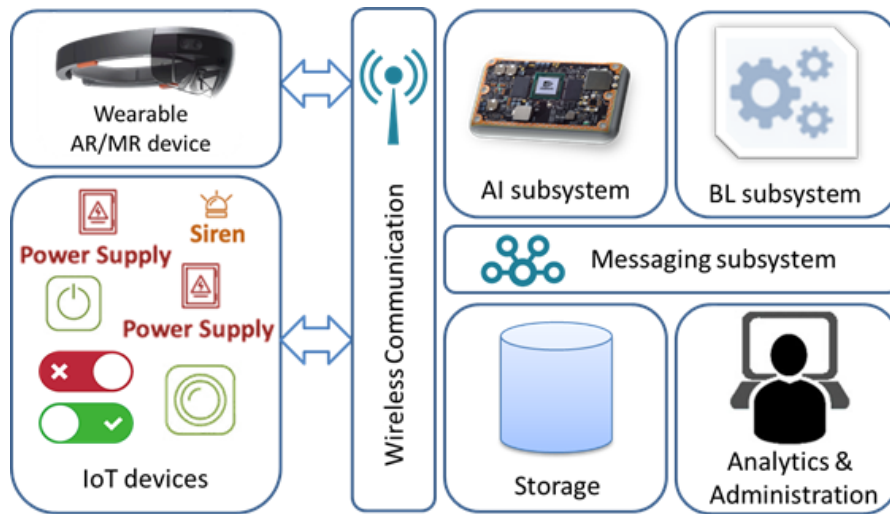


Figure 5: Architettura del sistema ENIGMA.

- **Motore di AI:** unità di elaborazione ad alte prestazioni, progettato per eseguire task di Intelligenza Artificiale. L'unità elabora i video provenienti dai dispositivi indossabili in tempo reale, e si occupa della localizzazione dell'operatore, del rilevamento e riconoscimento degli oggetti, dell'interazione uomo-oggetto e dell'individuazione del prossimo oggetto con cui avverrà l'interazione.
- **Business Logic:** le informazioni provenienti dal motore di Intelligenza Artificiale e dai dispositivi IoT vengono raccolte e combinate da questo modulo, che si occupa di prendere le decisioni relative al sollevamento degli allarmi relativi alla sicurezza ed al risparmio energetico.
- **Dispositivi IoT:** questi dispositivi forniscono lo stato di tutte le periferiche connesse al sistema ENIGMA, consentendone il controllo allo stesso tempo.

Altre componenti del sistema forniscono capacità di comunicazione (message broker), storage ad alta velocità di accesso (IMDB), storage persistente (DB non relazionali), interazione con l'esterno (API REST), gestionali (controllo remoto), e di analisi (web application).

4. Servizi

In questa sezione vengono riportati i servizi offerti dal sistema ENIGMA.

Realtà aumentata. Il sistema ENIGMA utilizza la Realtà Aumentata per fornire informazioni supplementari all'operatore, che può comunque mantenere libere le mani, garantendo la sicurezza. Allarmi, avvertimenti e notifiche vengono mostrati all'operatore sul visore olografico, e l'intera interfaccia può essere controllata vocalmente.

Localizzazione. Le informazioni sulla posizione

dell'operatore vengono utilizzate per mostrare suggerimenti in Realtà Aumentata, ed inoltre è attivo un logging che consente di visualizzare la posizione in tempo reale nella web application. Lo storico delle posizioni può essere utilizzato a fini di monitoraggio.

Rilevamento degli oggetti. Rilevare e riconoscere gli oggetti nell'ambiente circostante consente al sistema di selezionare e mostrare le informazioni appropriate all'operatore, e, grazie all'Intelligenza Artificiale ed al modulo di Business Logic, anche di estrarre statistiche quali il tempo di utilizzo o il carico di potenza relativo all'oggetto rilevato.

Interazione uomo-oggetto. Questo servizio consente di verificare la correttezza delle procedure che l'operatore sta eseguendo, ed anche di implementare politiche di risparmio energetico e prevenzione dell'usura in maniera automatica.

Next-Active Object. Attraverso questa funzionalità, attualmente ancora in sviluppo, sarà possibile prevenire situazioni rischiose, come ad esempio toccare elementi alimentati elettricamente. Il sistema, in tali casi, mostrerà degli allarmi e interagirà in maniera opportuna con il sottosistema IoT per prevenire il rischio.

Visual Analytics. La web application mette a disposizione una vista sinottica del laboratorio, in cui vengono riportati tutti i dispositivi ed il loro stato corrente (consumo di energia, picchi, allarmi rilevati ecc.). Inoltre viene fornito un log degli eventi rilevati.

Risparmio energetico. Questo servizio sfrutta in maniera combinata i sottosistemi di Intelligenza Artificiale e IoT per rilevare quando un dispositivo non è più utilizzato. In questo caso, il dispositivo interessato viene disalimentato per diminuire i consumi, o, se è presente un operatore in loco, viene mostrata una notifica che lo invita a provvedere.

5. Sicurezza

Il sistema è in grado di rilevare e risolvere correttamente i seguenti rischi per la sicurezza del lavoratore.

- **Rischio elettrico:** il lavoro con componenti elettroniche espone il lavoratore a rischio di folgorazione, derivante da un non avvenuto distacco dall'alimentazione o dalla carica residua conservata nei condensatori. Il sistema è in grado di avvertire tempestivamente l'operatore attraverso segnali visuali ed acustici qualora questo rischio sia presente, e allo stesso tempo di provvedere autonomamente a togliere alimentazione ai dispositivi interessati.
- **Pausa:** le norme per la tutela dei lavoratori prescrivono pause a intervalli regolari. Il sistema

ENIGMA può monitorare il tempo di lavoro continuo, suggerendo all'operatore quando è il momento di fermarsi e riposare.

- **Procedure di sicurezza:** attrezzi e dispositivi specialistici richiedono procedure specifiche. Il sistema fornisce consigli e suggerimenti, e controlla che tali procedure siano rispettate, incrementando la sicurezza dei lavoratori.

6. Conclusioni

Il sistema ENIGMA è un assistente indossabile finalizzato al miglioramento della sicurezza nei luoghi di lavoro ed al risparmio energetico. L'analisi del contesto in cui il lavoratore opera ed il rilevamento delle interazioni con gli oggetti hanno mostrato, nei test sperimentali, ottimi risultati, aprendo una strada promettente verso gli obiettivi successivi, quale l'anticipazione del prossimo oggetto attivo.

References

- [1] D. Damen, T. Leelasawassuk, O. Haines, A. Calway, W. W. Mayol-Cuevas, You-do, i-learn: Discovering task relevant objects and their modes of interaction from multi-user egocentric video., in: BMVC, volume 2, 2014, p. 3.
- [2] G. M. Farinella, G. Signorello, S. Battiato, A. Furnari, F. Ragusa, R. Leonardi, E. Ragusa, E. Scuderi, A. Lopes, L. Santo, et al., Vedi: Vision exploitation for data interpretation, in: Image Analysis and Processing—ICIAAP 2019: 20th International Conference, Trento, Italy, September 9–13, 2019, Proceedings, Part II 20, Springer, 2019, pp. 753–763.
- [3] R. Cucchiara, A. Del Bimbo, Visions for augmented cultural heritage experience, IEEE MultiMedia 21 (2014) 74–82.
- [4] S. Colombo, Y. Lim, F. Casalegno, Deep vision shield: Assessing the use of hmd and wearable sensors in a smart safety device, in: Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments, 2019, pp. 402–410.
- [5] F. Ragusa, A. Furnari, S. Livatino, G. M. Farinella, The meccano dataset: Understanding human-object interactions from egocentric videos in an industrial-like domain, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021, pp. 1569–1578.
- [6] F. Ragusa, A. Furnari, G. M. Farinella, Meccano: A multimodal egocentric dataset for humans behavior understanding in the industrial-like domain, arXiv preprint arXiv:2209.08691 (2022).

- [7] F. Osti, R. de Amicis, C. A. Sanchez, A. B. Tilt, E. Prather, A. Liverani, A VR training system for learning and skills development for construction workers, *Virtual Reality* 25 (2021) 523–538.
- [8] S. R. Sorko, M. Brunnhofer, Potentials of augmented reality in training, *Procedia Manufacturing* 31 (2019) 85–90.
- [9] P. Gurevich, J. Lanir, B. Cohen, R. Stone, Teleadvisor: a versatile augmented reality tool for remote assistance, in: *Proceedings of the SIGCHI conference on human factors in computing systems*, 2012, pp. 619–622.
- [10] M. Rebol, C. Hood, C. Ranniger, A. Rutenberg, N. Sikka, E. M. Horan, C. Gütl, K. Pietroszek, Remote assistance with mixed reality for procedural tasks, in: *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, IEEE, 2021, pp. 653–654.
- [11] L. Sun, H. A. Osman, J. Lang, An augmented reality online assistance platform for repair tasks, *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 17 (2021) 1–23.