

## INTERFACCIA CYBERFISICA UOMO-MACCHINA BASATA SU DEEP LEARNING

C. Nuzzi<sup>(1)</sup>, S. Pasinetti<sup>(1)</sup>, F. Docchio<sup>(1)</sup>, G. Sansoni<sup>(1)</sup>

<sup>(1)</sup>Dip. di Ingegneria Meccanica e Industriale, Università degli Studi di Brescia

mail autore di riferimento: c.nuzzi@unibs.it

### 1. INTRODUZIONE

Il paradigma **Industria 4.0** introduce il concetto di "smart factory", ovvero l'interconnessione tra sistemi informatici e sistemi fisici (detti sistemi *cyberfisici*). In particolare, questo lavoro di ricerca si focalizza sul concetto di "smart production", basato su una concreta collaborazione tra macchine, strumenti ed operatori per mezzo di supporti informatici [1]. In questo senso, le **workstation meta-collaborative** risultano essere di particolare interesse, perché a differenza delle *workstation collaborative* classiche lo spazio di lavoro dell'operatore e del robot non è necessariamente condiviso e la cooperazione può avvenire a prescindere dalla presenza di barriere fisiche tra i due. In questo lavoro di ricerca si prosegue l'indagine presentata in [2] [3] riguardante le potenzialità dei sistemi intelligenti basati su Deep Learning come strumento di interfaccia cyberfisica tra umano e macchina, inserendo un modello allenato ad hoc su dei gesti prescelti all'interno di una interfaccia di comando basata su ROS [4].

### 2. SISTEMA DI INTERFACCIA PROPOSTO

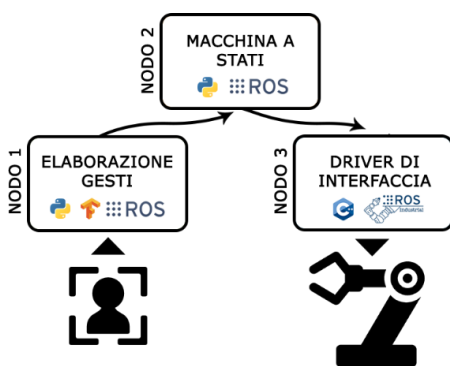


Fig. 1 Schema a blocchi del sistema proposto.

La scena e in particolare l'operatore vengono ripresi da alcune telecamere. I frame acquisiti dalla telecamera preposta all'acquisizione dei gesti vengono analizzati dal **nodo di elaborazione gesti** in modo da estrarre il comando dato dall'operatore e tradurlo come un codice operativo per il sistema. Questo codice verrà inviato al **nodo della macchina a stati** che si occuperà di gestire a livello software l'operazione selezionata dall'utente e creerà il pacchetto dati operativo da inviare al **nodo del driver di interfaccia** tra ROS e il controllore del robot. Quest'ultimo è fondamentale per interfacciarsi al software proprietario del robot che si andrà ad utilizzare e, a seconda della casa produttrice della macchina, è necessario utilizzare un driver differente.

### 3. RICONOSCIMENTO GESTI

A valle di uno studio sul linguaggio dei gesti per creare un linguaggio intuitivo e veloce da utilizzare per gli esseri umani, sono stati scelti 15 gesti statici per comporre le istruzioni del sistema (Fig. 2). Ogni gesto a una mano può essere eseguito sia con la mano destra che con la mano sinistra mostrando il palmo alla telecamera, per un totale di 2 varianti. Il gesto "Horiz" è l'unico che presenta 4 varianti: con il palmo rivolto verso la telecamera e con il dorso rivolto verso la telecamera, sia per la mano destra che per la mano sinistra.

Il modello che si è scelto di utilizzare è un Object Detector RFCN pre-allenato su COCO [5]. Una prima analisi delle performance ottenute dal

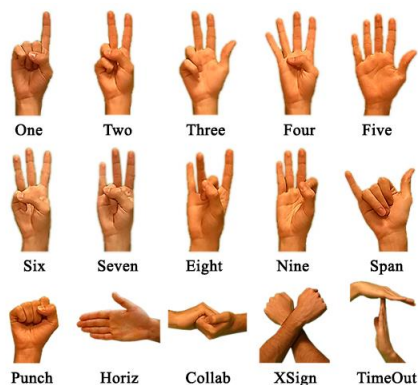


Fig. 2 Gesti proposti per il dataset sperimentale.

modello allenato su un dataset acquisito sperimentalmente ha evidenziato diverse difficoltà nel riconoscere i gesti numerici dal 6 al 9, soprattutto in presenza di molti oggetti diversi sullo sfondo. Questo ci ha portati a pensare che l'informazione di colore RGB non fosse sufficiente per risolvere il problema del riconoscimento anche in situazioni difficili, come potrebbero essere quelle reali di un ambiente industriale, e che fosse necessario aggiungere una o più modalità di acquisizione per migliorare le prestazioni.

#### 4. CONCLUSIONE E SVILUPPI FUTURI

Il lavoro di ricerca è ancora in corso, in particolare per definire il modello migliore da utilizzare all'interno del prototipo di interfaccia realizzato. Nello specifico, è attualmente in corso una fase di analisi e valutazione delle performance di numerose varianti dello stesso modello. In particolare, i modelli attualmente in fase di test comprendono l'utilizzo di un quarto canale corrispondente alla depth map della scena e/o l'utilizzo di uno spazio colore differente dal classico RGB, quali lo spazio **HSV** (Hue – Saturation – Value) [6] e lo spazio **TSL** (Tint – Saturation – Lightness) [7], quest'ultimo ritenuto particolarmente adatto per segmentare il colore della pelle rispetto al resto dell'immagine. Al termine di questa campagna di prove, si prevede di mettere a punto il modello migliore e valutare quindi le performance dell'intera interfaccia cyberfisica in situazioni di reale utilizzo, allestendo una o più stazioni meta-collaborative sperimentali.

#### RIFERIMENTI BIBLIOGRAFICI

- [1] M. Hermann, T. Pentek e B. Otto, «Design Principles for Industrie 4.0 Scenarios,» in *2016 49th Hawaii International Conference on System Sciences (HICSS)*, 2016.
- [2] C. Nuzzi, S. Pasinetti, M. Lancini, F. Docchio e G. Sansoni, «Deep Learning based Machine Vision: first steps towards a hand gesture recognition set up for Collaborative Robots,» in *Proceedings of the 2018 IEEE International Workshop on Metrology for Industry 4.0 and IoT*, Brescia, Italy, 2018.
- [3] C. Nuzzi, S. Pasinetti, M. Lancini, F. Docchio e G. Sansoni, «Deep Learning-based hand gesture recognition for Collaborative Robots,» *IEEE Instrumentation & Measurement Magazine*, vol. 22, n. 2, pp. 44-51, April 2019.
- [4] M. Quigley, B. P. Gerkey, K. Conley, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler e A. Ng, «ROS: an open-source Robot Operating System,» in *ICRA Workshop on Open Source Software*, 2009.
- [5] J. Dai, Y. Li, K. He e J. Sun, «R-FCN: Object Detection via Region-based Fully Convolutional Networks,» *arXiv preprint arXiv:1605.06409*, 2016.
- [6] A. R. Smith, «Color Gamut Transform Pairs,» *SIGGRAPH Computer Graphics*, pp. 12-19, 1978.
- [7] J. C. Terrillon, M. David e S. Akamatsu, «Automatic detection of human faces in natural scene images by use of a skin color model and of invariant moments,» in *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, 1998.