Riconoscimento di gesti mediante dispositivi a basso costo:
tecniche, applicazioni, prospettive

Vito Gentile, Salvatore Sorce, Alessio Malizia, Antonio Gentile

Sommario

Negli ultimi anni abbiamo assistito ad una grande diffusione dei cosiddetti "Kinect-like devices", ovvero dispositivi basati su un insieme di sensori a basso costo, che consentono di ottenere un'immagine di profondità della scena ripresa. L'alta accessibilità di questi dispositivi, principalmente in termini di costi, ne ha facilitato la diffusione nell'ambito del riconoscimento dei gesti in numerose applicazioni, sia commerciali che di ricerca. In questo articolo saranno inizialmente illustrati i principi generali su cui si fondano le principali tecniche utilizzate per riconoscere i gesti, sfruttando i dati ottenibili dai dispositivi "Kinect-like". Successivamente, saranno presentati alcuni ambiti applicativi, spaziando dal settore educativo-ricreativo a quello più scientifico (domotica, robotica ed ingegneria biomedica). In appendice, verranno elencati i principali prodotti disponibili in commercio e ne verrà presentata una sintetica analisi comparativa. Verrà inoltre descritto uno dei più noti e usati algoritmi di skeletal tracking, su cui si fonda la maggior parte delle soluzioni per il riconoscimento dei gesti.

Abstract

In the last decades, we have witnessed to the increase of the so-called Kinect-like devices, which are based on a set of low-cost sensors to acquire RGB and depth data of a scene. The high accessibility of such devices, mainly in terms of costs, has pushed their adoption as fundamental tool for gesture recognition in a large number of applications, both commercial and research-related ones. In this paper, we first discuss some of the general principles adopted by most of the main gesture recognition techniques described in literature. Then we present some application fields in which Kinect-like devices and gesture recognition algorithms have been used, ranging from educational-recreational examples to more complex

and scientific fields (e.g. domotics, robotics and biomedical engineering). In two annexes, we list and shortly compare the main features of the Kinect-like devices available on the market, and we describe one of the most popular algorithm for skeletal tracking, which is the basis for the gesture recognition.

Keywords: Gesture recognition; Kinect-like devices; Human-Computer Interaction; Touchless Interaction

1. Introduzione

Negli ultimi decenni, la ricerca nell'ambito dell'interazione uomo-macchina (o *Human-Computer Interaction*, *HCI*) ha visto un crescente interesse verso la scoperta di nuove modalità di interazione. L'innovazione e le nuove tecnologie, dalle più comuni (come i dispositivi mobili) a quelle meno diffuse, hanno contribuito in maniera significativa all'introduzione di nuove forme di interazione. Si è passato dall'interazione a riga di comando, al desktop, fino ad arrivare alle interfacce *touch*. Oggi, una delle nuove frontiere in termini di modalità di interazione sembra essere quella dell'interazione *touchless* e, in particolare, a gesti.

Per *interazione touchless* si intende una modalità di interazione che avviene senza che l'utente entri in contatto meccanico con alcun dispositivo o strumento fisico e tangibile del sistema [1]. Questo tipo di funzionalità può essere realizzata tramite l'uso di opportuni strumenti hardware, che forniscono dati utili ad un sistema informatico per "percepire" il comportamento interattivo di uno o più utenti. È importante notare che, secondo la precedente definizione di interazione touchless, interagire con un sistema utilizzando un controller, come accade ad esempio nell'interazione con Nintendo Wiimote o altri dispositivi simili (si veda ad esempio [2]), non è considerato una tipologia di interazione touchless. Al contrario, dispositivi che riconoscono la direzione dello sguardo (*eye tracker*), o i cosiddetti dispositivi *Kinect-like* [3], sono largamente accettati come validi esempi di dispositivi che permettono l'interazione touchless.

In questo articolo discuteremo dell'utilizzo di dispositivi *Kinect-like* al fine di implementare le principali tecniche utilizzate per il riconoscimento dei gesti, sfruttando i dati ottenibili da tali dispositivi. Verranno poi presentati alcuni ambiti applicativi, spaziando dal settore educativo-ricreativo a quello più scientifico (domotica, robotica ed ingegneria biomedica).

2. Riconoscimento di gesti statici

Secondo la definizione proposta da Henze et al. [4], un gesto *statico* è definito dalla *posizione* della mano nello spazio, mentre un gesto *dinamico* è definito dal *movimento* compiuto dalla mano. Ovviamente la definizione può essere estesa a gesti di altre parti o dell'intero corpo. In questa sezione si discuterà lo stato dell'arte relativo al riconoscimento dei gesti statici, con particolare attenzione a quelli della mano, per poi estendere la discussione al caso dei gesti corporali.

Nella sezione successiva verranno trattati gli approcci tipici per il riconoscimento di gesti dinamici.

2.1. Gesti statici della mano

Il processo di riconoscimento della posizione della mano è stato studiato da molti autori, e la letteratura fornisce numerose soluzioni. Spesso, soprattutto nei lavori più recenti, sono stati messi a punto sistemi in grado di funzionare in tempo reale sfruttando dispositivi *Kinect-like*.

In linea di massima, gli algoritmi di riconoscimento dei gesti statici della mano presentano alcuni tratti comuni. La Figura 1 mostra un tipico diagramma di flusso.

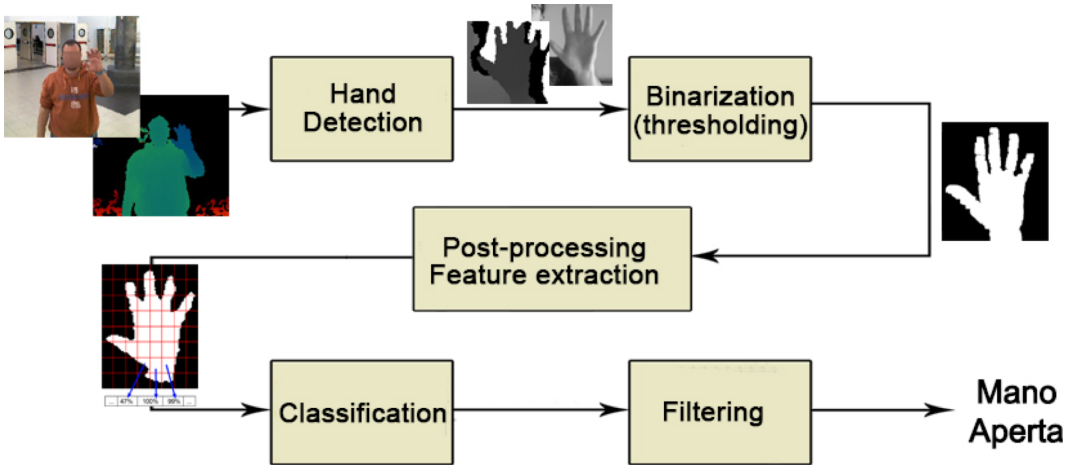


Figura 1.
Tipico diagramma di flusso per un algoritmo di riconoscimento di gesti statici della mano

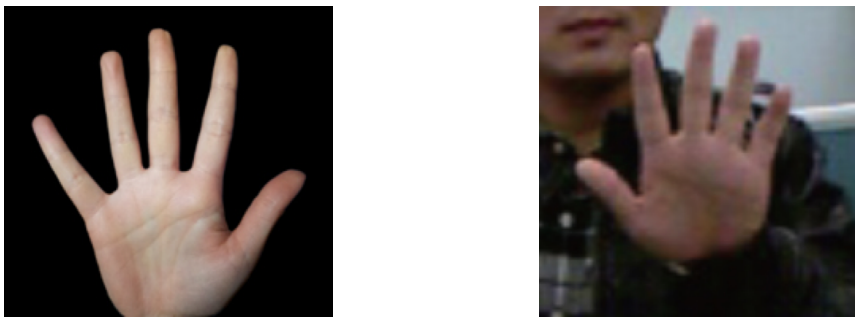
In alcuni casi il diagramma può non rispecchiare perfettamente gli algoritmi presenti in letteratura, in quanto le operazioni di *feature extraction* (estrazione di caratteristiche) possono essere basate anche su dati diversi dalle maschere binarie (come ad esempio i bordi o qualsiasi altra informazione significativa). La discussione seguente si baserà comunque su questo schema, che è quello seguito da molti dei lavori presi in esame.

2.1.1. Riconoscimento e localizzazione della mano (*hand detection*)

Il processo di *hand detection* consente l'individuazione della mano e della sua posizione all'interno di un'immagine comunque complessa. Il risultato di questo processo è una porzione dell'immagine (detta *region of interest*, *ROI*), contenente la mano. Sulla *ROI* estratta saranno effettuati tutti i successivi passi per l'estrazione delle caratteristiche ed il filtraggio.

La complessità del processo di localizzazione dipende da fattori quali la complessità della scena ripresa, nonché la disponibilità di dati aggiuntivi relativi

alla posizione della mano. La figura 2a mostra un caso favorevole, in cui la mano si trova in primo piano, su uno sfondo uniforme. In questo caso l'estrazione è semplice, essendo sufficiente procedere con il ritaglio dell'immagine (operazione conosciuta anche come *cropping*) basato sul calcolo del *bounding box* della mano (ovvero il minimo rettangolo contenente il contorno della mano). In casi meno controllati, come quello mostrato in figura 2b, un approccio di questo tipo può risultare più complesso.



(a) (b)

Figura 2.
Alcuni esempi di frames a livelli di difficoltà crescente per la localizzazione della mano

Talvolta è possibile utilizzare algoritmi di rimozione di uno sfondo statico (*background removal*) da una sequenza di immagini con elementi dinamici in primo piano (come la mano) [5]. In tal modo è successivamente possibile procedere al calcolo della ROI come precedentemente descritto. Tali algoritmi di *background removal* (che, da soli, possono funzionare in real-time) hanno delle limitazioni relative al fatto che, se in primo piano non vi è soltanto la mano, altre porzioni del corpo potrebbero essere erroneamente incluse nella ROI.

Utilizzando i dispositivi *Kinect-like*, l'approccio seguito in [6] si riconduce all'utilizzo dei giunti scheletrici, che possono essere agevolmente estratti tramite un algoritmo molto efficiente di *skeletal tracking* (si veda il relativo riquadro di approfondimento) [7]. Prendendo a riferimento il giunto relativo alla mano, si può centrare su di esso un rettangolo opportunamente dimensionato, che rappresenterà la ROI. Il dimensionamento di esso può essere valutato sulla base di considerazioni antropometriche, fissando euristicamente le sue dimensioni, oppure valutandole sulla base di quelle di altre parti del corpo.

2.1.2. Binarizzazione

Una volta individuata la ROI su cui operare per l'estrazione delle caratteristiche, si può procedere con la successiva fase, che è quella della *binarizzazione* (figura 3): si tratta di un'operazione che ha lo scopo di trasformare l'immagine della mano in una immagine binaria, che utilizza cioè due soli colori (tipicamente bianco e nero). Per questo processo, in generale, si possono utilizzare approcci basati sul colore della pelle, sui dati di profondità, o su entrambi questi tipi di informazioni.

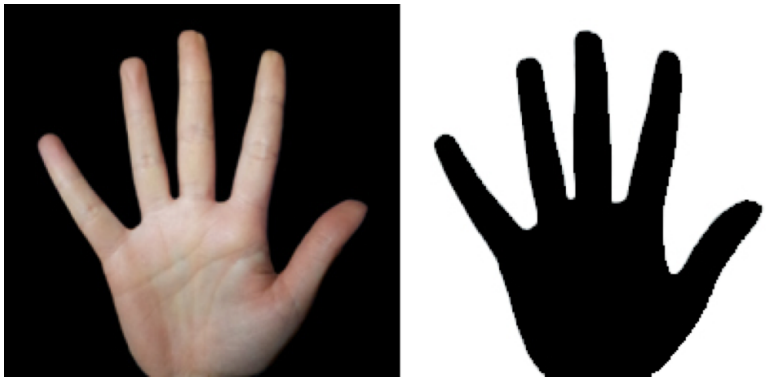


Figura 3.
Un esempio di binarizzazione di una mano

Alcuni autori hanno utilizzato approcci basati sul colore della pelle [8] [9], finalizzati all'estrazione della mano. A tal fine, sono stati proposti metodi che si basano sull'utilizzo di spazi di colore più appropriati di RGB, uniti a modelli probabilistici che assegnano ad ogni pixel una probabilità di appartenere alla pelle, in funzione del colore. In casi come quello di figura 2b, però, ciò può facilmente comportare errori, dal momento che anche altre zone del corpo vengono riconosciute, nonostante non facenti parte della mano.

Altri autori, invece, si basano sulla segmentazione basata unicamente sui dati di profondità [6] [10] [11]. Ciò ha il grande pregio di consentire un funzionamento indipendente dalle condizioni di illuminazione, ed è per questo motivo che è stato largamente utilizzato.

Si può migliorare il processo di binarizzazione, correggendo la maschera di profondità estratta, utilizzando le informazioni di colore fornite dalla videocamera RGB di un dispositivo *Kinect-like*. Approcci ibridi di questo tipo sono stati usati da alcuni ricercatori per ottenere i risultati più accurati da questa fase del processo [12].

La tabella 1 riassume le caratteristiche principali dei metodi di binarizzazione sopra accennati.

Metodo di binarizzazione	**Indipendente dalle variazioni di illuminazione**	**Robusto all'utilizzo di guanti e accessori**
Segmentazione basata sul colore della pelle	NO	NO
Segmentazione basata sulla mappa di profondità	SI	SI
Segmentazione ibrida, (colore + profondità)	Solo alcuni metodi	Solo alcuni metodi

Tabella 1.
Caratteristiche dei metodi di binarizzazione

2.1.3. *Post-processing* ed estrazione di caratteristiche

Una volta concluso il processo di binarizzazione (ma quel che segue può essere applicato anche prima di questa fase), si può procedere con la correzione del profilo della forma ottenuta, tramite operazioni di *post-processing* che, ad esempio, ne correggano le imperfezioni. Un semplice esempio, mostrato in figura 4, è l'utilizzo di un filtro mediano per lo smussamento [13], mentre soluzioni più complesse prevedono l'utilizzo di tecniche per la rimozione dell'avambraccio [9] [14] [15], eventualmente incluso nel risultato della segmentazione.

Figura 4.
Esempio di post-processing effettuato applicando un filtro mediano all'immagine binaria di una mano.

Ottenuta, quindi, un'immagine binaria rappresentante la forma della mano, essa deve essere trasformata in una forma che ne evidenzi le caratteristiche essenziali per il riconoscimento, e che possa al tempo stesso essere utilizzata per la classificazione. A tale scopo, sono state adottate molte tecniche. Le più interessanti sono quelle in grado di offrire un elevato grado di invarianza alla scala ed alla rotazione (per motivi che saranno meglio chiariti più avanti).

Alcune tecniche proposte in letteratura codificano il contorno della mano in modo che esso venga rappresentato da una o più funzioni lineari [14]. Esistono poi altre tecniche, più o meno complesse, basate sul calcolo di caratteristiche che possano descrivere un contorno (per esempio la combinazione di grandezze come l'area, il perimetro, il numero di concavità, eccetera). Molte di queste tecniche sono riassunte e descritte in [16].

Approcci alternativi si basano sull'estrazione di caratteristiche locali, calcolate cioè su sottoregioni (eventualmente - ma non necessariamente - organizzate gerarchicamente) delle immagini binarie originariamente utilizzate [15] [10] (figura 5). Altri autori, utilizzando anche le informazioni di profondità e di colore, ricavano lo scheletro della mano e lo adoperano per il riconoscimento [8], oppure estraggono descrittori di forma tridimensionali da un modello 3D ricavato dai dati di profondità [17]. In [6] e [11], vengono estratti dei descrittori *SURF (Speeded Up Robust Features)*, utilizzando la maschera di profondità per evidenziare la sola ROI all'interno dell'immagine RGB. Van den Bergh e Van Gool [12] utilizzano un sistema di riconoscimento basato sulle *Haarlets*, descritte a loro volta in [18].

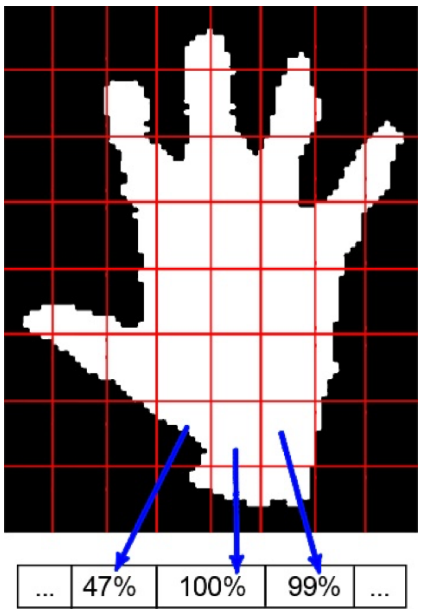


Figura 5.
Esempio di estrazione di caratteristiche locali, in cui si nota la suddivisone dell'immagine in sottoregioni

La tabella 2 mostra un confronto tra pro e contro dei metodi di post-processing presi in considerazione.

METODO	PRO	CONTRO
Descrittori di boundary	Buon livello di compressione Robusto a rotazione e scala	Richiede un alto livello di dettaglio dell'input, soprattutto per la differenziazione delle dita
Features locali	Si presta alla parallelizzazione Efficiente	Generico, non sempre fornisce una buona descrizione della forma della mano
Scheletro della mano	Garantisce un ottimo livello di dettaglio del riconoscimento	Richiede un alto livello di dettaglio dell'input, soprattutto per la differenziazione delle dita
Descrittori di forma 3D	Garantisce un ottimo livello di dettaglio del riconoscimento	Algoritmo di riconoscimento computazionalmente pesante
Descrittori SURF	Efficace per elaborazioni in tempo reale	Poco robusto alle variazioni di luminosità
2D e 3D Haarlets	Efficace per elaborazioni in tempo reale	Poco robusto alle variazioni di luminosità

Tabella 2.
Confronto tra tecniche di post-processing

2.1.4. Classificazione e filtraggio

Le tecniche più adoperate per la classificazione sono quelle basate sull'uso delle reti neurali [6] [19], o in alternativa quelle che utilizzano le macchine a vettore di supporto (*Support Vector Machine*, *SVM*) [11]. In entrambi i casi si tratta di strumenti matematici in grado di "apprendere" un algoritmo tramite una fase detta di "addestramento": utilizzando un insieme di input e output noti, una serie di algoritmi consente di "configurare" un programma in modo che esso "impari" ad ottenere un output valido anche per nuovi input.

Alcuni autori classificano i gesti statici seguendo anche approcci di tipo diverso. Ren et al. [14] si basano sul calcolo di una opportuna misura di distanza tra una rappresentazione della mano ed alcuni prototipi caratteristici di una classe. La classificazione avviene, quindi, in base alla minore misura di dissimilarità.

Subito dopo il processo di classificazione, talvolta è possibile applicare un filtro che rimuova il rumore dell'uscita. In [6], si utilizza un filtro di media temporale. In alternativa, è possibile modellare il rumore come gaussiano, ed adoperare filtri più complessi, come quello di Kalman [11].

2.2. Gesti statici del corpo

Abbiamo descritto finora alcuni principi su cui si basano le tecniche di riconoscimento di una "posa" della mano. È intuibile come molte delle idee sopra esposte possono essere utilizzate anche per il riconoscimento delle "pose" del corpo (che, secondo la definizione di Henze et al. precedentemente citata, sono anch'esse definite come gesti statici [4]). In letteratura, il processo di riconoscimento posturale si articola generalmente in due fasi:

1. estrazione delle caratteristiche;
2. classificazione vera e propria, per stabilire qual è la configurazione posturale in esame.

Diversi autori seguono lo schema appena descritto, utilizzando un algoritmo di *skeletal tracking* (si veda il relativo riquadro di approfondimento) per estrarre una serie di giunti scheletrici (che rappresentano le *caratteristiche* di cui sopra, in questo caso) da utilizzare per la classificazione [20] [21]. Gli autori utilizzano diverse metriche per valutare la distanza tra le posture esaminate e quelle appartenenti ad un set di riferimento, operando quindi una classificazione. È anche possibile riconoscere alcune nuove classi di posture, ed aggiungerle al set di riferimento (implementando ciò che viene definito un "sistema di apprendimento incrementale").

Altri lavori simili migliorano il riconoscimento tramite tecniche più o meno sofisticate mirate a risolvere i problemi di parziale occlusione del corpo [22].

È comunque possibile effettuare il riconoscimento posturale anche estraendo caratteristiche diverse dai giunti scheletrici (e quindi senza utilizzare alcun algoritmo di *skeletal tracking*). Biswas e Basu [10], ad esempio, utilizzano un approccio basato sulla suddivisione della sagoma del corpo (ricavata dalla binarizzazione dell'immagine di profondità), in sottoregioni organizzate gerarchicamente. Da ognuna di esse sono poi estratte delle caratteristiche locali, utilizzate tutte insieme per la classificazione. Quest'ultima avviene tramite l'uso di una *SVM* opportunamente addestrata.

3. Riconoscimento di gesti dinamici

In questa sezione si discuterà lo stato dell'arte relativo al riconoscimento dei gesti dinamici, costituiti cioè da una sequenza di gesti statici. Nella prima parte ci concentreremo sugli algoritmi utilizzati per il riconoscimento dei gesti dinamici della mano, per poi estendere la discussione al caso più ampio dei gesti del corpo. Il processo di riconoscimento dei gesti dinamici richiede l'analisi di più immagini consecutive, e di conseguenza è generalmente richiesto un carico computazionale più oneroso rispetto al riconoscimento dei gesti statici.

3.1. Panoramica

Per i gesti dinamici, siano essi relativi alla mano o al corpo intero, i lavori presenti in letteratura scientifica prevedono ancora un processo di estrazione di caratteristiche della parte del corpo interessata, utilizzabili poi in una successiva fase di classificazione. Nei casi dei gesti statici, queste caratteristiche sono tipicamente legate alla forma (che, per questo, sono identificate come *shape features*). Nei casi di gesti dinamici, anche grandezze come la velocità del movimento o la traiettoria di alcuni punti salienti diventano parametri di decisione rilevanti, che possono essere utilizzati per la classificazione. Per riuscire ad ottenere questo genere di informazioni occorre, quindi, tenere conto di finestre temporali successive, e dei vettori di caratteristiche relativi ad esse che vengono via via calcolati.

Quindi, un generico processo di riconoscimento dei gesti dinamici può essere schematizzato come mostrato in figura 6:

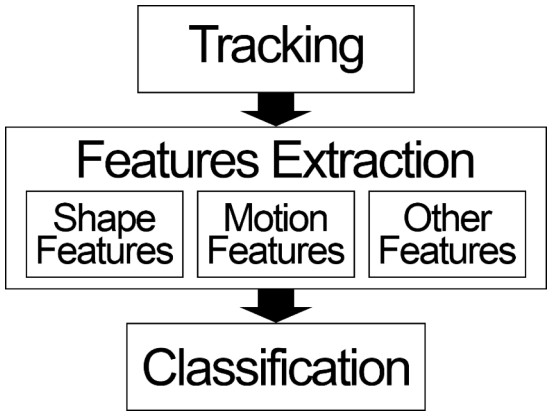


Figura 6.
Schema riepilogativo del processo di riconoscimento dei gesti dinamici.

Mentre per quanto riguarda le fasi di estrazione delle *shape features*, si può fare riferimento alle stesse tecniche viste per i gesti statici, le fasi di estrazione delle *motion features* (caratteristiche legate alle traiettorie dei punti, e quindi al movimento che definisce il gesto) richiedono considerazioni diverse. Per quanto riguarda la classificazione, le tecniche generalmente utilizzate sono quelle in grado di indirizzare dinamicamente il processo decisionale, in funzione dei

vettori di caratteristiche estratti di volta in volta da ogni immagine della sequenza che compone il gesto.

3.2. Gesti dinamici della mano

Nel riconoscere un gesto della mano, vi sono numerosi fattori da tenere in considerazione. In primis, durante il gesto, la postura delle singole dita può variare, e tale variazione può essere parte del gesto. Inoltre, la traiettoria tracciata dalla mano caratterizza anch'essa il gesto. In alcune situazioni, uno di questi due aspetti può essere ritenuto meno importante ai fini del riconoscimento, riducendo anche l'onere computazionale dell'intero processo. Xiao Jang et al. [23], per esempio, effettuano il solo riconoscimento della traiettoria, a patto però che la mano sia aperta: utilizzando un metodo di riconoscimento statico per riconoscere la corretta posa della mano, si può utilizzare la sola traiettoria come input per un classificatore basato su *Hidden Markovian Model* (*HMM*) (uno strumento matematico molto utilizzato in scenari come il riconoscimento dinamico di gesti).

Per gesti più complessi, in cui la configurazione delle singole dita ha una valenza significativa, è necessario tenere in considerazione anche le variazioni di forma della mano. In questo senso, il lavoro di Kurakin et al. [15] rappresenta una soluzione più completa. Essi, infatti, effettuano una serie di operazioni che estraggono diverse *shape features*. A queste, gli autori aggiungono la velocità del centro della mano (che è una *motion feature*), rappresentata dallo spostamento di essa tra due immagini successive. Queste caratteristiche vengono quindi classificate tramite uno strumento matematico chiamato *action graph* (definito in [24]).

Una delle principali applicazioni del riconoscimento dei gesti dinamici è quella del riconoscimento delle lingue dei segni. A tale scopo, Geetha et al. [25] utilizzano una tecnologia che consente di estrarre una serie di giunti scheletrici di entrambe le mani. Utilizzando sette punti dello scheletro di ogni mano, estratti da ogni frame, si possono tracciare le traiettorie che tali giunti delineano durante il movimento della mano. In questo modo si riassumono sia le informazioni sulla forma (date dai giunti scheletrici) sia quelle relative al moto.

Esistono comunque un gran numero di tecniche alternative a quelle citate, ma quelle citate rendono l'idea del tipo di approccio necessario per implementare il riconoscimento di gesti dinamici della mano.

3.3. Gesti dinamici del corpo

Sebbene, in linea di principio, il riconoscimento dei gesti dinamici del corpo non differisca molto da quello dei gesti dinamici della sola mano, la disponibilità di algoritmi di *skeletal tracking* come quelli già citati ne rende meno complicato lo sviluppo nelle applicazioni pratiche. Molti metodi, infatti, si basano semplicemente sull'uso di sequenze di posture [20] [21] [22], ciascuna delle quali non è altro che un insieme di giunti scheletrici che vanno a comporre il gesto dinamico del corpo. Tali metodi, inoltre, sono in grado di funzionare in tempo reale.

Ciò che differenzia i metodi non è tanto, quindi, l'estrazione delle *shape features* (che sono proprio i giunti scheletrici), quanto piuttosto l'utilizzo di *motion features* e le diverse possibilità di classificazione. Quest'ultimo processo può utilizzare diversi strumenti matematici (*Hidden Markovian Model*, reti neurali, *SVM*, reti bayesiane), la cui descrizione richiederebbe però un approfondimento notevole, ed esula dagli scopi del presente articolo.

4. Applicazioni

Le tecniche di riconoscimento dei gesti sono state utilizzate in svariati ambiti di ricerca, nonché in numerose applicazioni. Esse si prestano innanzitutto alla costruzione di interfacce multimodali, offrendo la possibilità di incrementare il livello di immersività dell'interazione tra l'utente ed il computer. Inoltre, riconoscere i gesti e le posture può rivelarsi molto utile nell'ambito della sicurezza degli utenti all'interno di ambienti domotici, al fine di prevenire incidenti domestici o segnalare malori.

Si possono definire alcune macroaree di interesse, all'interno delle quali trovano applicazione alcune tecniche di riconoscimento dei gesti (siano essi statici o dinamici). Esse sono:

- *area ludico-educativa*, che rappresenta il principale ambito di applicazione di molti dispositivi *Kinect-like*, in particolare Microsoft Kinect (nato per offrire modalità di interazione a gesti e innovative agli utenti di Xbox). La sfera educativa può essere inclusa in questa macroarea, dal momento che utilizzando la multimodalità dell'interfaccia, si può accrescere l'interesse di studenti verso varie discipline, facilitando i compiti degli educatori;
- *area domotica*, in cui si sfruttano le tecniche di riconoscimento dei gesti al fine di garantire una maggiore sicurezza dell'utente, ma anche per aumentare le possibilità di controllo delle componenti *smart* della casa;
- *area divulgativa*, in cui si utilizza il riconoscimento dei gesti per aumentare l'*immersione* dell'utente all'interno di un ambiente, al fine di accrescerne l'interesse. È il caso delle installazioni museali interattive, ma anche di proposte pubblicitarie finalizzate ad aumentare il coinvolgimento dell'utente, o ad accrescerne la curiosità;
- *area robotica*, nella quale i dispositivi *Kinect-like* sono stati ampiamente utilizzati per facilitare il riconoscimento delle azioni degli utenti che interagiscono con i robot (nell'ambito che viene definito di *Human-Robot Interaction*, *HRI*);
- *area bio-medica*, nella quale si sfruttano le funzionalità dei dispositivi *Kinect-like* sia in ambito riabilitativo (ad esempio per valutare eventuali problemi nei movimenti di un paziente) sia a scopi meno critici (ad esempio per interagire a gesti con immagini medicali proiettate su un display).

Nel seguito saranno analizzate alcune delle soluzioni proposte in letteratura, evidenziandone i pregi e le limitazioni.

4.1. Applicazioni ludico/educative

L'ambito ludico è uno dei principali campi di applicazione di alcuni dispositivi *Kinect-like*. Si pensi, ad esempio, a Microsoft Kinect, sviluppato e commercializzato principalmente per essere utilizzato insieme alla console Microsoft Xbox (figura 7). Per questa piattaforma, infatti, sono stati sviluppati e pubblicati centinaia di giochi basati su interazione a gesti, tuttora in vendita ed in continua evoluzione.

Figura 7.
Due utenti interagiscono a gesti con un gioco per Xbox tramite Microsoft Kinect.

Tuttavia, l'utilizzo di Microsoft Kinect in congiunzione con Xbox non è l'unico caso di applicazione ludico-educativa che sfrutta dispositivi *Kinect-like*. Infatti, è facile immagine come le possibilità offerte da questa tecnologia possano essere estese ed applicate nelle più svariate piattaforme. Questa idea è esattamente ciò che sta alla base di KinectEDucation [26], una community che ha lo scopo di promuovere l'utilizzo di Microsoft Kinect in ambito educativo, offrendo materiale didattico direttamente a studenti ed educatori, ma anche supporto per gli sviluppatori interessati in questo senso. Molti dei software proposti da KinectEDucation consentono agli studenti di imparare nozioni di base, utilizzando i gesti per creare un'interfaccia innovativa, che ne stimoli l'interesse. Per citare qualche esempio, sono stati sviluppati software per facilitare l'apprendimento della lingua dei segni, per esplorare ed imparare l'anatomia umana osservandola direttamente sul proprio corpo, nonché per ispirare la creatività tramite la possibilità di tracciare disegni o manipolare e muovere oggetti.

Un interessante lavoro mirato, invece, a facilitare il compito di un educatore, è il sistema proposto da Shuai Zhang et al. [27], che prevede l'integrazione dei dispositivi *Kinect-like* con alcune videoproiezioni per creare una lavagna virtuale proiettabile, senza l'utilizzo di alcuno schermo. Utilizzando un controller wireless (basato su tecnologia IR) delle dimensioni di una penna, si può infatti tracciare

un qualsiasi disegno, proiettarlo su una parete e visualizzarlo tramite video proiezione. Il tracciamento può essere realizzato effettuando il tracciamento della mano ripresa, con i metodi già descritti in precedenza.

Infine, alcune applicazioni possono essere quelle relative alla valutazione di particolari attività, legate ad alcuni sport come il golf, in cui la postura ha una rilevanza fondamentale. Lichao Zhang et al. [28] hanno studiato un sistema in grado di valutare la correttezza dei movimenti di un giocatore di golf, tramite l'assegnazione di un punteggio ad ogni prova di lancio della pallina da golf effettuata utilizzando un'apposita mazza.

4.2. Domotica

Negli ambiti della domotica e della home automation, i dispositivi *Kinect-like* sono stati utilizzati (e continuano ad essere studiati) sia per aumentare le possibilità di interazione con la casa ed i suoi componenti, utilizzando direttamente i gesti, sia per garantire un maggiore livello di sicurezza, tramite funzionalità di riconoscimento automatico di incidenti domestici (come descritto in [29]).

Relativamente al controllo domestico, ha particolare rilevanza il progetto chiamato WorldKit [30], che utilizza dispositivi *Kinect-like* congiuntamente ad opportune videoproiezioni per implementare funzionalità di controllo (figura 8). Sostanzialmente, questo sistema consente di trasformare le pareti di una casa in superfici con le quali è possibile interagire tramite gesti. Le possibilità offerte sono moltissime: si va dal controllo di svariati apparecchi (volume dello stereo o della televisione, impianto di climatizzazione, ecc.), fino alla pianificazione delle attività tramite proiezioni interattive.

Figura 8.
Un esempio di interazione supportata da WorldKit [30], in cui è visibile l'integrazione tra dispositivi Kinect-like e videoproiezioni.

In [31] vengono analizzate diverse applicazioni domotiche, sia per il controllo dell'ambiente guidato dai gesti, sia per la sicurezza degli utenti. Tra quelle studiate dagli autori, se ne possono citare alcune abbastanza interessanti.

Hands-Up [32], ad esempio, è un sistema che combina l'uso di un proiettore e di Microsoft Kinect per sfruttare il soffitto al fine di creare uno schermo col quale interfacciarsi tramite i gesti. Lo stesso tipo di funzionalità sono implementate anche in [33].

Un'altra applicazione che sfrutta Microsoft Kinect per l'utilizzo in ambito domestico è *Kinect in the Kitchen* [34], che utilizza il riconoscimento dei gesti in ambienti come la cucina, in cui l'interazione gestuale può consentire l'interazione con uno schermo pur avendo le mani sporche o unte. Gli autori affermano di aver avuto feedback molto positivi dagli utenti che hanno provato questo sistema.

4.3. Servizi Informativi

Le applicazioni in questo campo sfruttano i gesti non solo per l'interazione, ma soprattutto per accrescere l'interesse degli utenti alle informazioni da esse fornite.

Un tipico esempio è *Interactive Wall* [35], un progetto che ha consentito la realizzazione di un intero muro interattivo lungo più di 10 metri, animato in funzione dei gesti e degli spostamenti degli utenti che passavano davanti ad esso. In questo modo, è stato possibile applicare le funzionalità dei dispositivi *Kinect-like*, e le relative tecniche di riconoscimento dei gesti, al fine di stimolare l'interesse dei passanti a scopi pubblicitari. Progetti analoghi sono descritti in [36] e [37].

Anche in molte installazioni museali i dispositivi *Kinect-like* hanno trovato applicazione, risultando strumenti estremamente utili a stimolare l'interesse degli utenti verso la cultura [38] [39].

4.4. Robotica

Il riconoscimento dei gesti si presta non solo all'interazione tra uomo e computer, ma anche a contesti come la *Human-Robot Interaction* (*HRI*) e la *Robot-Robot Interaction* (*RRI*). È per questo che sono state spesso integrate diverse tecniche di riconoscimento dei gesti in sistemi robotici [40], ed anche i dispositivi *Kinect-like* sono stati utilizzati a tale scopo [41]. Un esempio delle recenti applicazioni di più alto livello è *JediBot* [41], un braccio robotico in grado di analizzare i gesti del corpo di un "avversario" dotato di spada (o bastone) per generare una risposta in tempo reale (figura 9). Altro tipo di applicazione, invece, è quella proposta in [42], dove gli autori discutono come sia possibile controllare uno scheletro robotico mediante un dispositivo *Kinect-like*.

4.5. Bio-medicina

Recentemente, i dispositivi *Kinect-like* sono stati utilizzati nel campo dell'ingegneria biomedica come parte di sistemi per il monitoraggio degli esercizi di riabilitazione dei pazienti. È il caso, ad esempio, di *Kinerehab* [43], un sistema per la riabilitazione fisica di giovani e adulti con disabilità motorie. In questo caso, il dispositivo *Kinect-like* è utilizzato per verificare la qualità del movimento, in modo da fornire un supporto al terapista sul trattamento da

effettuare. In maniera concettualmente analoga, Lange et al. [44] hanno sviluppato un gioco con interfaccia a gesti, compatibile con Microsoft Kinect, i cui scopi vengono raggiunti solo tramite movimenti finalizzati al trattamento riabilitativo. Alvarez e Grogan [45] hanno utilizzato questo tipo di dispositivi su pazienti affetti da morbo di Parkinson per migliorare anormalità motorie e disfunzioni dell'andatura.

Figura 9.
JediBot in azione.

Altre applicazioni biomediche utilizzano i dispositivi *Kinect-like* per il controllo della respirazione [46], nonché (più semplicemente) per consentire l'interazione a gesti con immagini medicali (molto utile in contesti operatori, in cui il medico è contemporaneamente impegnato nell'operare un paziente) [47].

5. Discussione

Il riconoscimento di gesti del corpo o di sue parti gioca oggi un ruolo chiave nell'ambito dell'interazione uomo-macchina. Infatti, la grande diffusione di sistemi informativi e la crescente richiesta di interattività, pongono nuove sfide circa le modalità di interazione, a causa di vincoli che non consentono l'uso dei consueti dispositivi di input.

Fino ad ora, la necessità di nuove modalità di interazione, e la loro applicazione, si è evidenziata soprattutto in ambiti molto particolari, in cui l'uso di dispositivi classici di input è scomodo o impossibile (telemedicina, controllo remoto di robot in zone ad alto rischio, ecc.), in ambito ludico (console per videogiochi), nell'*home entertainment*. Oggi, grazie alle nuove tecnologie, sono sempre più frequenti i casi in cui nuovi modi di interazione vengono produttivamente e utilmente impiegati.

Quelli descritti in questo articolo sono solo alcuni esempi di contesti in cui la capacità di interagire a distanza, mediante gesti in modalità *touchless*, ha aperto nuove prospettive e possibilità di impiego. Da tale apertura discende una

serie di sfide trasversali, dal punto di vista delle tecnologie, del fattore umano, delle possibili applicazioni.

Per quanto riguarda le tecnologie, l'analisi condotta in questo articolo dimostra come ormai sia economicamente accessibile e tecnicamente fattibile la realizzazione di sistemi di riconoscimento affidabile e veloce di gesti a scopo interattivo.

Nonostante l'accessibilità tecnica ed economica, tale modalità di interazione pone alcuni problemi relativamente al fattore umano, per esempio in termini di:

- *accettabilità sociale*: non è detto che le persone siano disposte a gesticolare (soprattutto in pubblico) davanti a un display per accedere alle informazioni da esso fornite;
- *naturalezza percepita*: i gesti da utilizzare per interagire dovrebbero essere "naturali" [48], nel senso di non necessitare di nessuna forma di addestramento all'uso, e risultare quindi massimamente intuitivi;
- *utilità effettiva*: l'utente deve percepire il vantaggio di tale tipo di interazione, a fronte di un servizio utile e innovativo;
- *efficacia*: i gesti devono condurre all'informazione desiderata in pochi passaggi.

In conclusione, i vincoli posti dall'interazione gestuale in termini di sviluppo, messa in opera e validazione, hanno finora limitato la sua adozione sui sistemi informativi di largo uso. La natura *touchless* di tale interazione, e le sue conseguenti potenzialità applicative, costituiscono però il motore principale della spinta evolutiva in corso, che è dimostrata dal vasto interesse in ambito scientifico. Le soluzioni in fase di studio e la disponibilità di tecnologie abilitanti a costi accessibili, lasciano presagire che, in un futuro non lontano, l'interazione gestuale uomo-macchina verrà sempre più adottata e riconosciuta come "naturale", alla stregua dell'interazione tra umani.

6. Bibliografia

[1] R. de la Barré, P. Chojecki, U. Leiner, L. Mühlbach e D. Ruschin, «Touchless Interaction-Novel Chances and Challenges,» in *Human-Computer Interaction. Novel Interaction Methods and Techniques*, Jacko, Julie A., 2009, pp. 161-169.

[2] A. Bellucci, A. Malizia, P. Diaz e I. Aedo, «Don't touch me: multi-user annotations on a map in large display environments,» in *Proceedings of the International Conference on Advanced Visual Interfaces (AVI '10)*, 2010.

[3] V. Gentile, S. Sorce e A. Gentile, «Continuous Hand Openness Detection Using a Kinect-Like Device,» in *2014 Eighth International Conference on Complex, Intelligent and Software Intensive Systems (CISIS)*, Birmingham, UK, 2014.

[4] N. Henze, A. Löcken, S. Boll, T. Hesselmann e M. Pielot, «Free-hand gestures for music playback: deriving gestures with a user-centred

process,» in *9th International Conference on Mobile and Ubiquitous Multimedia*, New York, NY, USA, 2010.

[5] J. Sun, W. Zhang, X. Tang e H.-Y. Shum, «Background cut,» in *Proceedings of the 9th European conference on Computer Vision (ECCV'06)*, Graz, Austria, 2006.

[6] S. Sorce, V. Gentile e A. Gentile, «Real-time Hand Pose Recognition Based on a Neural Network Using Microsoft Kinect,» in *Proceedings of the Eighth International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA 2013)*, Compiegne, France, 2013.

[7] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman e A. Blake, «Real-time human pose recognition in parts from single depth images,» in *2011 IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, USA, 2011.

[8] H. Matos, H. P. Oliveira e F. Magalhães, «Hand-Geometry Based Recognition System A Non Restricted Acquisition Approach,» in *9th International Conference on Image Analysis and Recognition (ICIAR)*, Aveiro, Portugal, 2012.

[9] M. K. Bhuyan, R. N. Debanga e K. K. Mithun, «Fingertip Detection for Hand Pose Recognition,» *Internation Journal on Computer Science and Engineering (IJCSE),* vol. 4, n. 3, 2012.

[10] K. K. Biswas e K. B. Saurav, «Gesture Recognition using Microsoft Kinect®,» in *5th International Conference on Automation, Robotics and Applications (ICARA)*, Wellington, New Zealand, 2011.

[11] A. D. Bagdanov, A. Del Bimbo, L. Seidenari e L. Usai, «Real-time hand status recognition from RGB-D imagery,» in *21st International Conference on Pattern Recognition*, Tsukuba, Japan, 2012.

[12] M. Van den Bergh e L. Van Gool, «Combining RGB and ToF cameras for real-time 3D hand gesture interaction,» in *IEEE Workshop on Applications of Computer Vision (WACV)*, Kona, HI, USA, 2011.

[13] S. Marchand-Maillet e Y. M. Sharaiha, Binary Digital Image Processing: A Discrete Approach, Academic Press, 2000.

[14] Z. Ren, J. Yuan e Z. Zhang, «Robust Hand Gesture Recognition Based on Finger-Earth Mover's Distance with a Commodity Depth Camera,» in *Proceedings of the 19th ACM international conference on Multimedia*, New York, NY, USA, 2011.

[15] A. Kurakin, Z. Zhang e Z. Liu, «A real time system for dynamic hand gesture recognition with a depth sensor,» in *20th European Signal Processing Conference (EUSIPCO)*, Bucharest, Romania, 2012.

[16] Y. Mingqiang, K. Kidiyo e R. Joseph, «A survey of shape feature extraction techniques,» *Pattern Recognition,* pp. 43 - 90, November 2008.

[17] S. Poonam, S. Anbumani e M. Dinesh, «Dynamic Hand Pose Recognition using Depth Data,» in *International Conference on Pattern Recognition*, Istanbul, Turkey, 2010.

[18] M. Van den Bergh, E. Koller-Meier e L. Van Gool, «Real-Time Body Pose Recognition Using 2D or 3D Haarlets,» *International Journal of Computer Vision,* vol. 83, n. 1, pp. 72 - 84, June 2009.

[19] G. P. Zhang, «Neural Network for Classification: A Survey,» *IEEE Transactions on Systems, Man, and Cybernetics, Part C,* vol. 30, n. 4, pp. 451-462, Novembre 2000.

[20] S. Monir, S. Rubya e H. S. Ferdous, «Rotation and scale invariant posture recognition using Microsoft Kinect skeletal tracking feature,» in *12th International Conference on Intelligent Systems Design and Applications (ISDA)*, Kochi, India, 2012.

[21] F. D. Zainordin, H. Y. Lee, N. A. Sani, Y. M. Wong e C. S. Chan, «Human pose recognition using Kinect and rule-based system,» in *World Automation Congress (WAC)*, Puerto Vallarta, Mexico, 2012.

[22] H. P. H. Shum, E. S. L. Ho, Y. Jiang e S. Takagi, «Real-Time Posture Reconstruction for Microsoft Kinect,» *IEEE Transactions on Cybernetics,* vol. 43, n. 5, pp. 1357 - 1369, 2013.

[23] X. Jiang, X. Lu, L. Chen, L. Zhou e S. Shen, «A Dynamic Gesture Recognition Method Based on Computer Vision,» in *6th International Congress on Image and Signal Processing (CISP 2013)*, Hangzhou, China, 2013.

[24] W. Li, Z. Zhang e Z. Liu, «Expandable data-driven graphical modeling of human actions based on salient postures,» *IEEE Trans. Circuits Syst. Video Techn.,* vol. 18, n. 11, p. 1499 – 1510, 2008.

[25] M. Geetha, C. Manjusha, P. Unnikrishnan e R. Harikrishnan, «A vision based dynamic gesture recognition of Indian Sign Language on Kinect based depth images,» in *International Conference on Emerging Trends in Communication, Control, Signal Processing & Computing Applications (C2SPCA)*, Bangalore, India, 2013.

[26] J. Kissko, December 2011. [Online]. Available: http://www.kinecteducation.com/blog/2011/11/15/what-is-kinecteducation-all-about/.

[27] S. Zhang, W. He, Q. Yu e X. Zheng, «Low-cost interactive whiteboard using the Kinect,» in *International Conference on Image Analysis and Signal Processing (IASP)*, Hangzhou, China, 2012.

[28] L. Zhang, J.-C. Hsieh, T.-T. Ting, Y.-C. Huang, Y.-C. Ho e L.-K. Ku, «A Kinect based Golf Swing Score and Grade System using GMM and SVM,» in *5th International Congress on Image and Signal Processing (CISP)*, Chongqing, Sichuan, China, 2012.

[29] H. Haggag, M. Hossny, S. Haggag, S. Nahavandi e D. Creighton, «Safety applications using Kinect technology,» in *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, San Diego, CA, USA, 2014.

[30] R. Xiao, C. Harrison e S. E. Hudson, «WorldKit: Rapid and Easy Creation of Ad-hoc Interactive Applications on Everyday Surfaces,» in *31st Annual SIGCHI Conference on Human Factors in Computing Systems*, Paris, France, 2013.

[31] A. C. d. C. Correia, L. C. d. Miranda e H. H. Hornung, «Gesture-Based Interaction in Domotic Environments: State of the Art and HCI Framework Inspired by the Diversity,» *INTERACT 2,* vol. 8118, pp. 300-317, 2013.

[32] J. Oh, Y. Jung, Y. Cho, C. Hahm, H. Sin e J. Lee, «Hands-up: motion recognition using kinect and a ceiling to improve the convenience of human life,» in *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, Austin, Texas, USA, 2012.

[33] H.-J. Kim, K.-H. Jeong, S.-K. Kim e T.-D. Han, «Ambient Wall: Smart Wall Display interface which can be controlled by simple gesture for smart home,» in *SIGGRAPH Asia 2011 Sketches*, Hong Kong, 2011.

[34] G. Panger, «Kinect in the Kitchen: Testing Depth Camera Interactions in Practical Home Environments,» in *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, Austin, Texas, USA, 2012.

[35] Flightphase, May 2014. [Online]. Available: http://www.flightphase.com/expanded-media/interactive-wall-at-ud.

[36] H. Stindl, June 2011. [Online]. Available: http://www.horizont.at/home/detail/digital-station-branding-premium-inszenierung.html.

[37] Kinect Hacks, December 2011. [Online]. Available: http://www.kinecthacks.com/kinect-used-in-university-campaign/.

[38] C.-K. Hsieh, W.-C. Liao, M.-C. Yu e Y.-P. Hung, «Interacting with the past: Creating a time perception journey experience using kinect-based breath detection and deterioration and recovery simulation technologies,» *Journal on Computing and Cultural Heritage (JOCCH),* vol. 7, n. 1, 2014.

[39] C.-S. Wang, D.-J. Chiang e Y.-C. Wei, «Intuitional 3D Museum Navigation System Using Kinect,» in *Information Technology Convergence*, Springer Netherlands, 2013, pp. 587-596.

[40] J. Triesch e C. v. d. Malsburg, «Robotic gesture recognition,» in *Gesture and Sign Language in Human-Computer Interaction*, Springer Berlin Heidelberg, 1998, pp. 233-244.

[41] T. Kröger, K. Oslund, T. Jenkins, D. Torczynski, N. Hippenmeyer, R. B. Rusu e O. Khatib, «JediBot – Experiments in Human-Robot Sword-Fighting,» *Springer Tracts in Advanced Robotics,* vol. 88, pp. 155-166, 2012.

[42] J. Ekelmann e B. Butka, «Kinect Controlled Electro-Mechanical Skeleton,» in *Proceedings of IEEE Southeastcon*, Orlando, FL, USA, 2012.

[43] Y.-J. Chang, S.-F. Chen e J.-D. Huang, «A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities,» *Research in Developmental Disabilities,* vol. 32, n. 6, pp. 2566-2570, November–December 2011.

[44] B. Lange, C.-Y. Chang, E. Suma, B. Newman, A. Rizzo e M. Bolas, «Development and evaluation of low cost game-based balance rehabilitation tool using the microsoft kinect sensor,» in *Annual*

International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, Boston, MA, USA, 2011.

[45] M. Alvarez e P. Grogan, «Connecting with Kinect® To Improve Motor and Gait Function in Parkinson Disease,» *Neurology,* vol. 78, April 2012.

[46] M. Alnowami, B. Alnwaimi, F. Tahavori, M. Copland e K. Wells, «A quantitative assessment of using the Kinect for Xbox360 for respiratory surface motion tracking,» in *SPIE Medical Imaging*, 2012.

[47] G. C. S. Ruppert, L. O. Reis, P. H. J. Amorim, T. F. d. Moraes e J. V. L. d. Silva, «Touchless gesture user interface for interactive image visualization in urological surgery,» *World Journal of Urology,* vol. 30, n. 5, pp. 687-691, 2012.

[48] A. Malizia e A. Bellucci, «The artificiality of natural user interfaces,» *Communications of ACM,* vol. 55, n. 3, pp. 36-38, Marzo 2012.

Biografia

Vito Gentile è studente di dottorato in Ingegneria dell'Innovazione Tecnologica presso il Dipartimento di Ingegneria Chimica, Gestionale, Informatica, Meccanica (DICGIM) dell'Università degli Studi di Palermo. I temi di ricerca di cui si occupa sono quelli dell'interazione uomo-macchina e dell'ubiquitous computing, con particolare interesse verso l'interazione touchless a gesti con i display pubblici.

Ha studiato Ingegneria Informatica presso l'Università degli Studi di Palermo, dove ha ottenuto la laurea magistrale col massimo dei voti e la lode nel 2013.

Email: vito.gentile@unipa.it

twitter: @ViGentile

Salvatore Sorce è ricercatore presso il Computer-Human Interaction Laboratory (CHILab) del Dipartimento di Ingegneria Chimica, Gestionale, Informatica. Meccanica (DICGIM) dell'Università degli Studi di Palermo. È inoltre co-fondatore e project manager di InformAmuse S.r.l., spin-off accademico dell'Università degli Studi di Palermo. Attualmente si occupa di ubiquitous computing e sistemi pervasivi; interazione uomo-macchina; display pubblici; dispositivi indossabili; programmazione mobile. Salvatore è inoltre senior member di ACM.

Ha studiato Ingegneria Informatica all'Università degli Studi di Palermo, dove si è laureato nel 2001 e dove ha ottenuto il dottorato di ricerca in Ingegneria Informatica nel 2006.

Email: salvatore.sorce@unipa.it.

Twitter: @salvosorce

Alessio Malizia è Senior Lecturer in Human-Computer Interaction presso il Dipartimento di Computer Science della Brunel University London (UK) e membro dell'Istituto di Human-Centred Design. Si è trasferito a Brunel dall'Universidad Carlos III de Madrid (Spagna), dove era Professore Associato in interazione uomo-macchina e social computing. Ha lavorato in precedenza all'Università La Sapienza di Roma, nonché presso IBM, SGI e Xerox PARC, presso il gruppo di Human-Document Interaction. Alessio è inoltre distinguished speaker e senior member di ACM e membro di IEEE.

Email: alessio.malizia@brunel.ac.uk

Twitter: @cikay72

Antonio Gentile è professore associato presso l'INovative Computer Architecture Laboratory (INCA Lab) del Dipartimento di Ingegneria Chimica, Gestionale, Informatica, Meccanica (DICGIM) dell'Università degli Studi di Palermo. Appassionato imprenditore, dal 2009 è anche CEO e fondatore di InformAmuse S.r.l., spin-off accademico dell'Università degli Studi di Palermo, nonché presidente e fondatore di Jujo Inc, una startup americana fondata nel 2013. Il prof. Gentile ha conseguito il titolo di Ph.D. in ingegneria elettrica ed informatica presso il Georgia Institute of Technology, Atlanta (USA) nel 2000. Attualmente si occupa di sistemi portabili di elaborazione ad alto rendimento, architetture per l'elaborazione di immagini e video, sistemi ed architetture embedded, riconoscimento del parlato, interfacce uomo-macchina e mobile computing. È inoltre associate editor di *Integration* e dell'*International Journal of Grid and Utility Computing*. Antonio è inoltre senior member di IEEE ed ACM, nonché membro di AEIT.

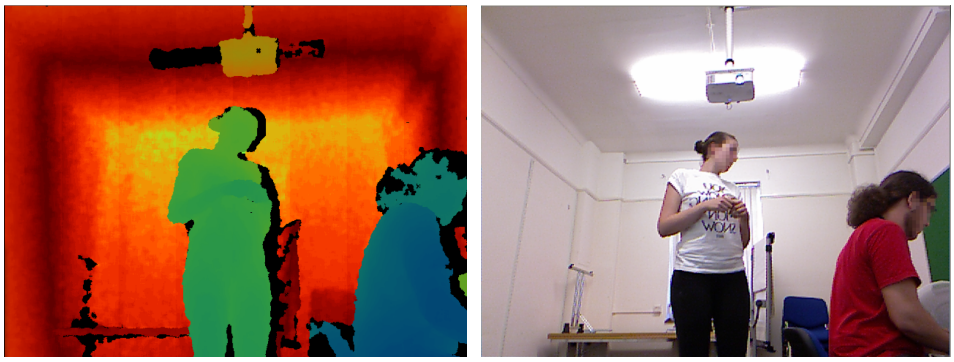
Email: antonio.gentile@unipa.it.

Twitter: @tonneiro

Approfondimento 1: I dispositivi Kinect-like

I dispositivi *Kinect-like* sono costituiti da un insieme di sensori a basso costo, e possono essere riassunti dalle seguenti caratteristiche:

- consentono di acquisire una rappresentazione in 3D della scena ripresa in tempo reale (figura 1a);
- consentono di acquisire una rappresentazione RGB della scena ripresa in tempo reale (figura 1b);
- sono poco costosi (dell'ordine di poche centinaia di euro).



(a) (b)

Figura 1.
Un tipico esempio dei dati RGB-D ottenibili tramite un dispositivo Kinect-like

Nella maggior parte dei dispositivi *Kinect-like* disponibili in commercio, la rappresentazione 3D della scena viene realizzata usando un sistema di raggi infrarossi (IR), costituito da un emettitore e da un sensore. L'emettitore proietta nell'ambiente esaminato un pattern ben definito di spot luminosi la cui disposizione è nota. La ricostruzione avviene confrontando la distribuzione dei punti proiettati sulla scena attuale con il pattern di riferimento così come dovrebbe essere visto dal sensore se esso fosse proiettato su una superficie posta ad una distanza ben definita e perfettamente parallela al piano della coppia emettitore-sensore. Come esempio, si può pensare al pattern di riferimento come ad una coperta a pois distesa su un piano e posta ad una distanza ben definita (figura 2a). Se un oggetto entra a contatto con la coperta (figura 2b), questa verrà deformata (figura 2c), e di conseguenza si potrà dedurre la disposizione spaziale dell'oggetto sulla base della nuova disposizione dei pois della coperta.

Un'alternativa al metodo appena descritto, che in genere è più costosa ma che Microsoft ha reso disponibile a basso costo con l'introduzione della versione 2 di Kinect, è il calcolo della profondità misurando il tempo di volo dei raggi infrarossi inviati dall'emettitore. Ciò consente, in genere, di ottenere dati di profondità più accurati.

I dispositivi *Kinect-like* attualmente disponibili sul mercato sono molti. Alcuni produttori hanno rilasciato diverse versioni di uno stesso dispositivo, talvolta anche abbastanza diverse l'una dall'altra in termini di funzionalità, SDK e hardware. Di seguito vedremo un confronto tra alcuni di questi dispositivi.

(a) (b) (c)

Figura 2.
Inserendo un oggetto (a) dietro una coperta a pois (b), la disposizione dei pois verrà deformata (c), e tale deformazione può consentire di risalire alla forma dell'oggetto. In maniera analoga, il pattern di raggi infrarossi proiettato e deformato dagli oggetti della scena consente di risalire alle informazioni di profondità.

Microsoft Kinect

Prodotto originariamente destinato all'ambito ludico (insieme alla console Xbox 360), Microsoft Kinect fu messo in vendita nel Novembre 2010, mentre nel 2012 è stata rilasciata un versione leggermente perfezionata, ma destinata ad un utilizzo su Windows. Oggi fuori produzione, si può trovare ancora in vendita a prezzi che variano tra i 20 ed i 200 €.

Figura 3.
Microsoft Kinect, nella versione per Xbox360.

Depth camera					RGB Camera			Altri sensori	microfono	inclinazione variabile
risoluzione	profondità	frame rate	campo visivo	tecnologia	Risoluzione	Frame rate	Campo visivo			
320x240	Default: 0.8 - 4 m Near mode: 0.4 - 3 m	30 fps	58.5° x 45.6°	PrimeSense PS1080-A2	1280x960	30 fps	62.0° x 48.6°	Accelerometro a tre assi	Array di microfoni (stereo mic)	Sì, motorizzata. Campo visivo ± 27°

Tabella 1.
Specifiche tecniche di Microsoft Kinect.

Nel 2014 è stata rilasciata una seconda versione di Microsoft Kinect, compatibile sia con la console Xbox One che (tramite un apposito adattatore) utilizzabile su PC con Microsoft Windows 8. La nuova versione utilizza una depth camera basata sul calcolo del tempo di volo, e risulta molto più precisa della versione precedente. La versione per Xbox One, a Febbraio 2016, è in vendita a circa 150 €, cui va aggiunto il costo dell'adattatore che consente di collegare questo dispositivo alla porta USB di un PC

Figura 4.
Kinect per Xbox One

Depth camera					RGB Camera			Altri sensori	microfono	inclinazione variabile
risoluzione	profondità	frame rate	campo visivo	tecnologia	Risoluzione	Frame rate	Campo visivo			
512x424	0.4 - 4.5 m	30 fps	70.6° x 60.0°	Microsoft X871141-001	1920×1080	30 fps	84.1° x 53.8°	Accelerometro a tre assi	Array di microfoni (stereo mic)	Non presente

Tabella 2.
Specifiche tecniche di Microsoft Kinect v2.

Asus Xtion PRO Live

Un altro dispositivo che ha trovato numerose applicazioni nella ricerca è stato prodotto da Asus, in collaborazione con PrimeSense. Rilasciato nel 2012, Asus Xtion PRO Live è in vendita (a Maggio 2015) per circa 170 €. Asus Xtion PRO Live è stato ispirato da un altro dispositivo, prodotto inizialmente da PrimeSense stessa, chiamato *Carmine*.

Figura 5.
Asus Xtion PRO Live

Depth camera					RGB Camera			Altri sensori	microfono	inclinazione variabile
risoluzione	profondità	frame rate	campo visivo	tecnologia	Risoluzione	Frame rate	Campo visivo			
320x240	0.8 – 3.5 m	30 fps	58° x 45°	PrimeSense PS1080-A2	1280x1024	30 fps	58° x 45°	N.D.	N.D.	Sì, manuale

Tabella 3.
Specifiche tecniche di Asus Xtion.

Altri dispositivi simili

Oltre ai dispositivi *Kinect-like* appena citati, vi sono altre soluzioni degne di nota ma meno diffuse delle precedenti.

(a) (b) (c) (d)

Figura 6.
Leap Motion Controller (a), SoftKinetic DS325 (b), Intel RealSense 3D camera (c), Creative Senz3D(d).

Il *Leap Motion Controller* (LMC), sviluppato dalla Leap Motion Inc. (figura 6a), ha avuto una discreta fortuna nell'ambito della ricerca. Lo scopo principale di questo dispositivo è il riconoscimento e tracciamento di mani e dita, con un grado di precisione di circa 2 mm. Il LMC è in grado di funzionare a 200 fps, e di riconoscere un range di profondità che varia tra 25.4 mm e 1.2 m, sfruttando due camere a infrarossi e tre LED. LMC viene generalmente utilizzato in modo diverso dai dispositivi fin qui descritti, in quanto viene poggiato su un piano orizzontale ed il suo campo visivo si estende verticalmente. Sebbene questo dispositivo abbia molte analogie con le tecnologie descritte fin qui, esso non rientra nella categoria dei dispositivi *Kinect-like* in quanto non include una

camera RGB. Esso è comunque disponibile a basso costo (il prezzo di lancio, nel 2013, era di circa 80 dollari USA).

Un'alternativa ai dispositivi *Kinect-like* più diffusi è stata prodotta dalla compagnia belga *SoftKinetic*, al fine di offrire una tecnologia di visione 3D per i PC, l'elettronica portatile, le auto e le macchine in genere. Tale compagnia produce un sensore di profondità a tempo di volo, basato su una tecnologia brevettata. Questo tipo di sensore è utilizzato in due dispositivi prodotti da SoftKinetic, ovvero *DepthSense 325* (DS325 – figura 6b) e *DepthSense 311* (DS311), che includono quindi anche una camera RGB e sono disponibili a prezzi relativamente accessibili (il costo di DS311 è di 299 dollari USA; DS325 può essere invece acquistato per 250 dollari USA). SoftKinetic fornisce anche un SDK compatibile con Windows e Visual Basic, ma anche questo tipo di prodotti non hanno avuto il successo di Microsoft Kinect o Asus Xtion, probabilmente per via del minore potere di marketing di SoftKinetic.

Anche Intel ha recentemente iniziato a sviluppatore una tecnologia che consente l'interazione a gesti basata sui dati di profondità della scena, chiamata RealSense. Tale tecnologia è stata integrata in molti prodotti commerciali, tra cui i dispositivi *Intel RealSense 3D Camera* (figura 6c) e *Creative Senz3D* (figura 6d). Quest'ultimo include una camera RGB con risoluzione di 1280x768 ed un sensore di profondità in grado di funzionare a 30 fps, con un campo visivo di 74° ed un range di profondità tra i 15.24 ed i 99.05 centimetri. Dal momento che si tratta di una tecnologia relativamente giovane, Intel RealSense non è stata molto applicata nella ricerca, soprattutto se paragonata con Asus Xtion o Microsoft Kinect.

Approfondimento 2: Skeletal tracking

Una delle funzionalità più utilizzate quando si ha a che fare con i dispositivi *Kinect-like*, è la possibilità di ottenere una serie di giunti scheletrici a partire dall'immagine di profondità di un utente. Ciò è reso possibile tramite una classe di algoritmi nota con il nome di *skeletal tracking*. In particolare, uno dei più noti algoritmi di questo tipo (ed integrato in Microsoft Kinect SDK) è stato sviluppato da Microsoft Research.

L'algoritmo in esame è conosciuto come *pose recognition in parts* (riconoscimento della postura in parti), nome dovuto al modo in cui i giunti scheletrici vengono ottenuti. A partire da una singola immagine di profondità, infatti, la sagoma del corpo umano (ricavata dall'immagine di profondità) è suddivisa in sezioni, ciascuna rappresentante diverse parti del corpo (mani, testa, braccia, gambe, ecc.).

Dopo aver provveduto ad eliminare qualsiasi elemento di sfondo, estraendo la fisionomia dell'utente, viene operata un'etichettatura probabilistica sull'immagine, che assegna ad ogni punto del corpo dell'utente le probabilità che tale punto appartenga ad ogni porzione del corpo. In tal modo, verranno generate tante distribuzioni di probabilità quante sono le parti del corpo prese in esame. A questo punto, i giunti scheletrici altro non sono che le mode statistiche di ciascuna distribuzione di probabilità (figura 1).

Figura 1.
Processo di etichettatura dell'immagine di profondità, seguito dall'ottenimento dei giunti scheletrici del corpo umano ripreso.

L'etichettatura probabilistica di cui sopra viene effettuata tramite un processo di classificazione, che utilizza a sua volta un algoritmo basato sull'addestramento di alberi decisionali. Senza scendere nei dettagli, si precisa che tale algoritmo ha richiesto una fase di addestramento, in cui l'insieme di dati utilizzati per l'apprendimento è stato generato artificialmente con alcuni modelli 3D. Da tali modelli, infatti, può essere ricavata sia un'immagine di profondità (input), sia la corrispondente etichettatura probabilistica (output atteso), in situazioni tipiche e particolarmente significative per garantire un apprendimento quanto più

soddisfacente. In particolare, Microsoft ha ricavato un database di 500.000 immagini, dalle quali è stato ottenuto un insieme di addestramento composto da 10.000 posture.

Una volta ottenuta l'etichettatura (che può essere pensata come una *texture* da applicare ai modelli 3D, in cui ogni colore identifica una parte del corpo), si ricavano i singoli punti che rappresentano i giunti scheletrici. Essi vengono calcolati tramite un algoritmo di ricerca della moda di ciascuna distribuzione di probabilità (ognuna corrispondente ad una parte del corpo).

I ricercatori Microsoft che si sono occupati dello sviluppo dell'intero algoritmo di *skeletal tracking*, affermano che un'implementazione ottimizzata di questo algoritmo permette l'analisi di un'immagine e l'estrazione dei giunti scheletrici in circa 5 ms (utilizzando la GPU integrata in Xbox 360), sufficiente a garantire un funzionamento in tempo reale.