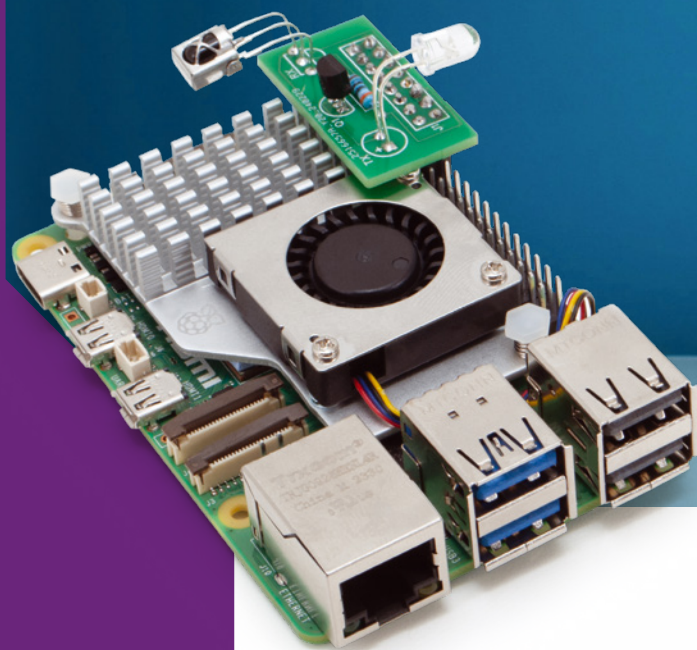


# TELECOMANDO UNIVERSALE AI MEDIAPIPE PER TV

di ALOISIO ANTONIO

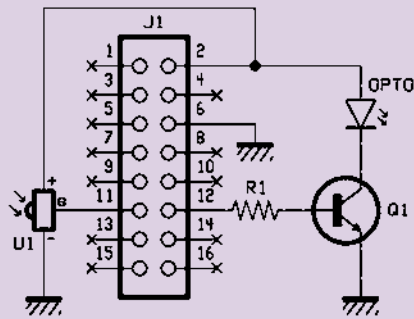


Realizziamo un dispositivo basato su Raspberry Pi capace di riconoscere gesti predefiniti della mano ed associarli a comandi ad infrarosso.

## H

MI, acronimo di Human-Machine Interface, è un termine che si riferisce a qualsiasi tecnologia o dispositivo capace di far interagire gli esseri umani con le macchine. Le HMI sono fondamentali in una vasta gamma di settori, dalla produzione industriale all'elettronica di consumo, e sono essenziali per semplificare e ottimizzare le interazioni tra gli utenti e le macchine. Le HMI più usate sono:

- **Interfacce utente grafiche (GUI)**, sono le più comuni e includono schermi grafici che consentono agli utenti di interagire con le macchine attraverso icone, pulsanti e menu, tipiche ad esempio dei sistemi operativi per pc.
- **Interfacce utente vocali (VUI)**, consentono agli utenti di interagire con le macchine attraverso comandi vocali.



Questo tipo di HMI sta diventando sempre più popolare con l'avvento dell'assistenza vocale come Siri di Apple, Google Assistant e Amazon Alexa.

- **Interfacce utente tattili:** consentono agli utenti di interagire con le macchine attraverso il tocco fisico, come schermi touchscreen, tastiere e mouse.
- **Interfacce utente gestuali:** consente agli utenti di interagire con le macchine attraverso movimenti fisici, come gesti delle mani rilevati da telecamere o sensori di movimento, che sarà il focus di questo progetto.

Il campo di applicazione spazia dall'automazione industriale (utilizzate per monitorare e controllare macchine e processi industriali), all'elettronica di consumo (i touchscreen sono comuni nei dispositivi consumer come smartphone, tablet, computer e televisori). Nei sistemi domotici, le HMI consentono agli utenti di controllare luci, termostati, dispositivi di sicurezza e altri dispositivi domestici intelligenti. Sono fondamentali nei veicoli per consentire agli utenti di interagire con sistemi di navigazione, sistemi di intrattenimento e altre funzionalità. L'interazione deve seguire dettami di ergonomia e funzionalità non sempre raggiunti in tutte le HMI, basti pensare alle recenti critiche da parte di Euro NCAP, nei confronti della complessità dei veicoli che hanno spostato controlli chiave sugli schermi centrali, obbligando i conducenti a distogliere lo sguardo dalla strada, aumentando il rischio di incidenti per distrazione (fonte Quattroruote). La realizzazione di HMI più intuitive (user-friendly) è quindi uno degli obiettivi che le aziende cercano sempre più di raggiungere. Quello dell'interazione con le macchine è quindi un ambito con cui ci confrontiamo quotidianamente ed è divenuto cruciale in una vasta gamma di settori, contribuendo a migliorare l'efficienza, la sicurezza e l'esperienza complessiva degli utenti. Oggi, molti dei sistemi HMI sfruttano l'intelligenza

artificiale (AI) per il loro funzionamento.

Il progetto che vi proporremo in questo articolo, userà il riconoscimento dei gesti della mano per dare dei comandi ad un televisore. Useremo un sistema di AI per sostituire uno dei dispositivi più comuni e "rassicuranti" nella vita di tutti i giorni, il telecomando. Alla fine del progetto sfidiamo chiunque a non divertirsi accendendo il televisore con "la forza Jedi" dell'apertura del palmo della mano. L'approccio del progetto è ludico, ma vuole essere una porta di ingresso sull'uso dei sistemi di intelligenza artificiale (AI) da parte di tutti in questo caso mirata alla creazione di HMI gestuale.

## SOFTWARE

L'intelligenza artificiale (AI) è un campo dell'informatica che si concentra sullo sviluppo di sistemi e tecnologie in grado di simulare i processi di intelligenza umana, cioè pensare e risolvere problemi a cui siamo abituati noi esseri umani.

Il "machine learning" (ML) è invece un sottoinsieme fondamentale della AI che studia come far sì che un sistema informatico possa imparare da **dati** passati per fare previsioni o prendere decisioni, senza essere esplicitamente programmati per farlo. Il primo passo nel ML è l'acquisizione dei dati che possono essere di vario tipo, come testo, immagini, suoni o numeri. Possono provenire da sensori, database o persino dai social media. L'importanza dei dati di qualità è fondamentale nel determinare il successo di un modello di ML.

Gli **algoritmi** di ML sono invece gli strumenti che permettono ai sistemi informatici di imparare dai dati. Esistono vari tipi di algoritmi, ognuno dei quali è adatto a compiti specifici. Ad esempio, gli algoritmi di regressione possono essere utilizzati per fare previsioni basate su dati storici, mentre gli algoritmi di classificazione sono utili per assegnare categorie ai dati.

Viene poi il processo di **addestramento** di un sistema (algoritmo + dato) di ML, cioè l'alimentazione dell'algoritmo con dati etichettati, cioè dati che hanno già delle risposte corrette associate. L'algoritmo utilizza quindi questi dati per imparare le relazioni. Questo passo richiede una certa quantità di tempo e risorse computazionali, ma è fondamentale per la creazione di modelli accurati. Una volta addestrato, il modello deve essere **valutato** per determinare quanto bene si comporta su dati in esso non presenti.

Questo passaggio è cruciale per capire se il modello è in grado di generalizzare bene oltre i dati con cui è stato addestrato.



Se il modello non soddisfa le prestazioni desiderate, può essere necessario **ottimizzarlo** regolando vari parametri o utilizzando tecniche di regolazione. Questo processo è noto come ottimizzazione del modello e può richiedere un po' di sperimentazione per trovare la combinazione ottimale di parametri.

## MEDIAPIPE

Questo grosso lavoro di creazione dell'algoritmo, addestramento ed ottimizzazione è fornito in modo Open Source da Google con la potentissima piattaforma MediaPipe.

Il significato del termine "MediaPipe" è legato al concetto di "pipeline dei media". Si potrebbe brutalmente tradurre il termine "pipeline" come condotta o tubatura, nel contesto informatico lo accostiamo al concetto di **flusso di dati** e del consequenziale loro processo. Abbiamo infatti detto che i dati sono i mattoni portanti dei sistemi di ML. Il termine media è nel nostro parlare quotidiano, quindi con esso identifichiamo il mezzo che gestisce questo flusso di dati.

MediaPipe viene definita da Google on-device, cioè anche capace di funzionare su un dispositivo hardware senza necessità di collegamento internet. Ciò rende i nostri progetti portatili ed integrabili in svariati contesti.

Google aggiunge il termine "no-code" alla sua piattaforma, cioè fornisce un sistema che necessita di poca conoscenza della programmazione informatica per raggiungere il suo obiettivo. Concetto che conosciamo bene con l'universo Arduino, che ha aperto il mondo dell'elettronica e della scrittura del codice informatico anche a chi non ne aveva le basi. Ci sentiamo di dire che anche MediaPipe mira in questa direzione, proponendo in modo semplice un argomento in origine molto complesso come l'intelligenza artificiale.

Vediamo subito le potenzialità andando all'indiriz-

zo internet: [https://mediapipe-studio.webapps.google.com/studio/demo/gesture\\_recognizer](https://mediapipe-studio.webapps.google.com/studio/demo/gesture_recognizer).

Occorre effettuare il login con un account Google. Diamo il consenso all'uso della web cam da parte del browser.

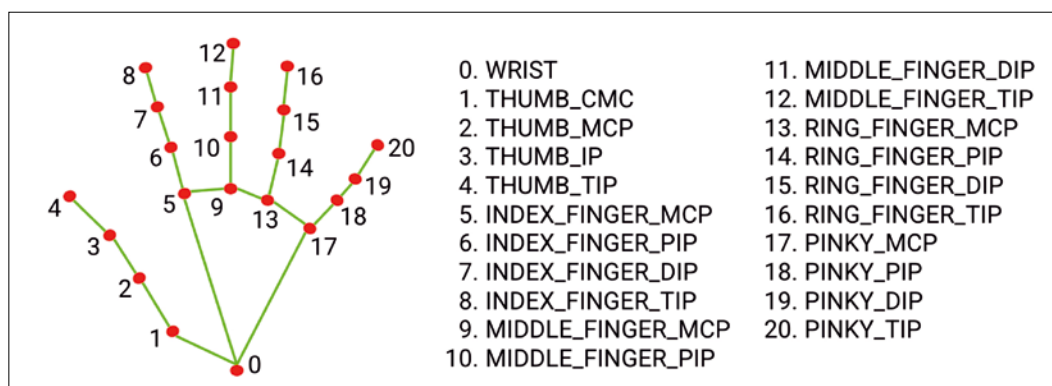
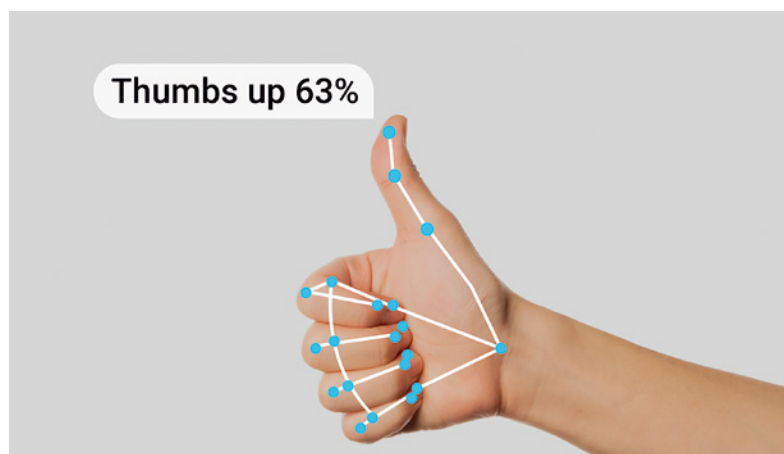
Inquadrando la nostra mano, destra o sinistra è indifferente, poiché identifica anche quale sia ed eseguiamo uno dei seguenti gesti:

- 1 – pugno chiuso
- 2 – palmo aperto
- 3 – indice che punta in su
- 4 – pollice in su
- 5 – pollice in giù
- 6 – simbolo della vittoria con due dita
- 7 – ti amo nella lingua dei gesti (posizione finale).

La pagina mostrerà, in tempo reale e sovrascritto al video, 21 punti chiave del palmo della mano collegati tra di loro mediante segmenti colorati **Fig.1**.

I punti chiave vengono definiti anche come **landmark** e rappresentano le parti significative delle mani, come le punte delle dita, i pollici, i palmi e così via. Il tracciamento preciso di questi landmark

**Fig.1**  
Esempio di riconoscimento del gesto pollice su.



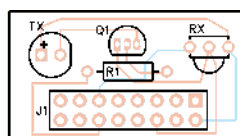
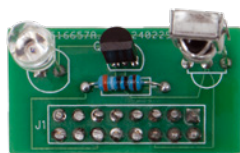
**Fig. 2**  
Punti riconosciuti sul palmo della mano - landmark.

## Elenco Componenti:

R1: 390 ohm  
Q1: 2N3904  
OPTO1: TSAL7600  
U1: IR38DM

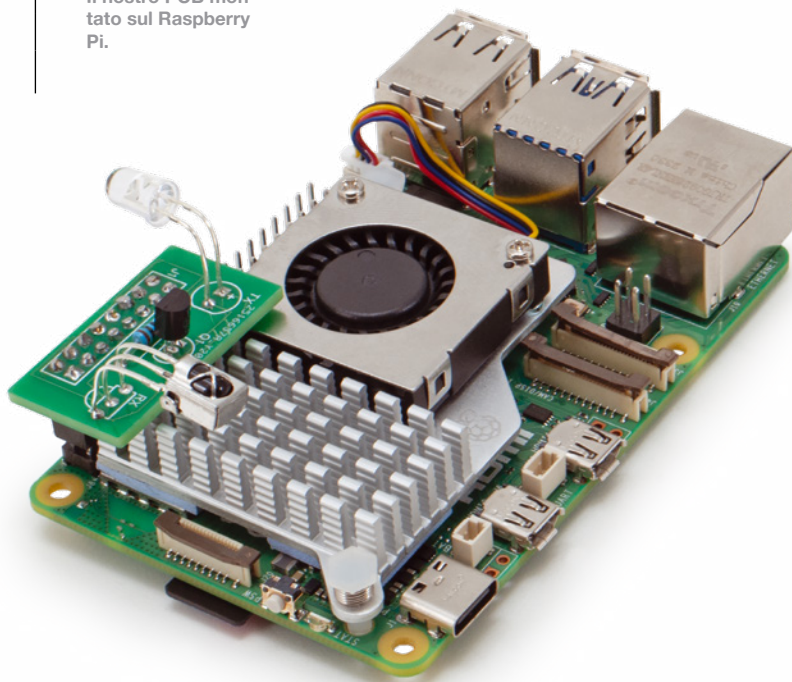
### Varie:

- Strip Femmina 2x8 vie passo 2.54 mm
- Circuito Stampato S1749 (17x31 mm)



consente di comprendere la posa e i movimenti delle mani nell'ambito di un'applicazione (**Fig.2**). Nella pagina web, sotto il riquadro del video, vi saranno anche una serie di informazioni aggiuntive, come la percentuale di sicurezza nel aver riconosciuto il gesto eseguito ed ovviamente il tipo di gesto. MediaPipe offre quindi una soluzione chiamata "Hand Tracking" (tracciamento delle mani) che consente il riconoscimento e il tracciamento delle mani in tempo reale all'interno di un flusso video. Questo modulo di MediaPipe è stato denominato di conseguenza "HandLandmarks". Il modulo inizia rilevando la presenza di mani nel frame video e identifica le regioni dell'immagine che contengono le mani. Una volta rilevate le mani, inizia a tracciare i landmark delle mani.

**Fig. 3**  
Il nostro PCB montato sul Raspberry Pi.



Con i landmark delle mani identificati, il modulo può stimare la posa delle mani all'interno del frame video. Questo include informazioni come l'orientamento delle mani, la direzione delle dita e altri dettagli che possono essere utilizzati per interpretarne i gesti.

MediaPipe è basata su TensorFlow lite, cioè la grande piattaforma di ML sempre di Google, ma ne differisce nel fatto che è già pre-addestrato.

Il riconoscimento gestuale sfrutta ad esempio circa 30000 immagini prese dal mondo reale.

Potrete anche addestrare direttamente voi il sistema, con immagini di altri gesti seguendo il tutorial

al seguente link: [https://developers.google.com/mediapipe/solutions/customization/gesture\\_recognizer](https://developers.google.com/mediapipe/solutions/customization/gesture_recognizer).

Divertitevi ad esplorare gli altri sistemi di ML forniti da MediaPipe come ad esempio la

classificazione delle immagini, rilevamento degli oggetti o dei volti. La documentazione ivi presente

vi permetterà di approfondire gli argomenti più di interesse. MediaPipe è anche multi piattaforma,

infatti si possono creare progetti per Android, iOS ed Web (JavaScript). Nel nostro progetto useremo

**Python** cioè uno dei linguaggi di programmazione più popolari nel campo del ML (e non solo) per

diverse ragioni: la sua semplicità d'uso, la sintassi intuitiva e la vasta gamma di risorse didattiche e

librerie disponibili lo rendono un'ottima scelta per i principianti e gli esperti di programmazione.

Un articolo non basterebbe neanche come premessa per imparare il grande mondo di TensorFlow, MediaPipe ed Python. Ma vorremo che vi

servisse per capirne le capacità e vi permettesse di indirizzare lo studio su argomenti indispensabili nel

campo dell'informatica presente e futura.

## MEDIAPIPE SU RASPBERRY PI

Entriamo allora nel pratico ed installiamo MediaPipe sul nostro Raspberry Pi 5, in modo da utilizzarlo subito on-device. Collegatevi alla vostra board

come meglio siete abituati (noi consigliamo VNC, per semplicità e potenza di utilizzo, troverete un

tutorial sul nostro sito <https://furanet.it/progetti/2023/04/20/realizzare-un-sistema-di-sorveglianza-con-raspberry-pi-e-opencv/>).

Python è nativamente presente nel sistema operativo di Raspberry Pi. Come sempre prendete i

diritti da amministratore con il comando `sudo bash` e accertatevi di avere tutto aggiornato con `apt update` ed `apt upgrade`.

Installiamo la libreria MediaPipe per Python con il comando da terminale:

```
pip install mediapipe
```

Mensile di elettronica applicata, attualità scientifica, novità tecnologiche.

# Elettronica In

[www.elettronica.in.it](http://www.elettronica.in.it)

oltre l'elettronica