



UNIVERSITÀ
POLITECNICA
DELLE MARCHE

FACOLTÀ DI INGEGNERIA

CORSO DI LAUREA TRIENNALE IN INGEGNERIA BIOMEDICA

Definizione di una procedura per la raccolta di immagini mediante un banco di fotosegnalamento

Definition of a procedure for collection of images by a work bench for photographic
identification

Candidate:

Fabrizia De Fabritiis

Relatore:

Prof.ssa Milena Martarelli

Anno Accademico 2021-2022

A Simone

Indice

Lista delle Figure	iii
Lista delle Tabelle.....	v
Introduzione.....	1
Scopo della Tesi.....	2
Struttura della Tesi.....	2
Capitolo 1 Stato dell'arte.....	3
1.1 Fotosegnalamento	3
1.2 Videosorveglianza	5
Capitolo 2 Materiali e metodi.....	9
2.1 Descrizione del banco	9
2.2 Descrizione telecamere	10
2.3 Descrizione braccio di posizionamento (meccanica e movimentazione)	14
2.4 Descrizione software di controllo movimentazione e acquisizione telecamere (LabView).....	16
2.5 Descrizione del software di acquisizione video (OBS studio).....	20
Capitolo 3 Sperimentazione	24
3.1 Protocollo acquisizione foto	24
3.2 Protocollo acquisizione video	25
3.3 Protocollo dichiarazione di consenso alla privacy	25
3.4 Acquisizione di un soggetto.....	27
Capitolo 4 Analisi dei risultati	35
Conclusioni.....	37
Sviluppi Futuri	37
Bibliografia.....	38
Sitografia.....	39

Lista delle Figure

Figura 1.1 - Esempio di un soggetto presente all'interno del database FERET [11].	5
Figura 1.2 - Sensori maggiormente utilizzati per la videosorveglianza e loro principali caratteristiche.....	6
Figura 1.3 – Immagine di una telecamera posta all'interno di un bancomat.....	7
Figura 1.4 – Immagine estrapolata dal video acquisito dalla telecamera 4 in laboratorio. ...	7
Figura 2.1 - Chip della <i>Raspberry Pi Zero W</i> utilizzata per l'acquisizione di immagini nel fotosegnalamento.....	11
Figura 2.2 - Custodia di protezione della <i>Raspberry Pi Zero W</i> utilizzata per proteggere i componenti interni.....	11
Figura 2.3 - Punti di vista sul piano verticale utilizzati per ottenere le foto segnaletiche...	12
Figura 2.4 - Struttura utilizzata per la riproduzione della videosorveglianza in laboratorio.	14
Figura 2.5 - Funzionamento schematizzato del braccio di posizionamento.....	15
Figura 2.6 - Immagine del front panel con evidenziati gli input.	17
Figura 2.7 - Immagine del front panel con evidenziati gli output.	18
Figura 2.8 - Immagine del block diagram che mostra il ciclo while per il controllo del motore che mette in rotazione il supporto delle telecamere.	19
Figura 2.9 - Immagine del block diagram che mostra il ciclo for per il controllo delle telecamere.....	19
Figura 2.10 – Elenco delle sorgenti per la cattura del video.	20
Figura 2.11 - Proprietà di una telecamera all'interno di OBS studio.	21
Figura 2.12 – Ordinamento delle telecamere.....	21
Figura 2.13 - Visualizzazione finale dopo l'applicazione del filtro “ <i>source record</i> ”.	22

Figura 2.14 – Struttura dei passaggi per selezionare la cartella dove salvare i video.	22
Figura 2.15 - Posizione del comando utile ad avviare la registrazione.	23
Figura 3.1 - Modulo della privacy consegnato ai soggetti che si sottoponevano alla procedura.	26
Figura 3.2 - Foto relative al primo angolo.	27
Figura 3.3 - Foto relative al secondo angolo.	27
Figura 3.4 - Foto relative al terzo angolo.	28
Figura 3.5 - Foto relative al quarto angolo.	28
Figura 3.6 – Foto scattata dalla telecamera frontale.	28
Figura 3.7 - Foto relative al quinto angolo.	29
Figura 3.8 - Foto relative al sesto angolo.	29
Figura 3.9 - Foto relative al settimo angolo.	29
Figura 3.10 – Output completo del fotosegnalamento.	30
Figura 3.11 – Esempio di un frame del video catturato dalla telecamera 1.	31
Figura 3.12 – Esempio di un frame del video catturato dalla telecamera 2.	31
Figura 3.13 – Esempio di un frame del video catturato dalla telecamera 3.	32
Figura 3.14 – Esempio di un frame del video catturato dalla telecamera 4.	33
Figura 3.15 – Esempio di un frame del video catturato dalla telecamera 5.	33
Figura 3.16 – Esempio di un frame del video catturato dalle 5 telecamere, il soggetto indossa degli occhiali da sole.	34

Lista delle Tabelle

Tabella 4.1 – Elenco di tutti i soggetti del dataset (X = completato).....	35
--	----

Introduzione

Il riconoscimento facciale è una delle tecniche più utilizzate per l'identificazione di un soggetto poiché sfrutta la biometrica naturale del volto. Per biometrica si intende l'approccio automatizzato per autenticare l'identità di una persona, utilizzando le caratteristiche fisiologiche uniche come: l'impronta digitale, il volto, la voce, la firma. Il suo punto di forza è che, essendo basata su tratti unici che sono parte dell'individuo, rende più difficile l'errore o la contraffazione. In questo modo, solamente un utente legittimo ha la possibilità di accedere ai servizi che consentono l'autenticazione biometrica [1]. Gli approcci iniziali al riconoscimento facciale erano basati su metodologie di *Computer Vision*. Queste tecniche ottenevano una buona accuratezza nei dataset, dove alcuni parametri come posa, luce ed espressioni erano fisse, ma erano insufficienti per estrarre features che fossero invarianti ai cambiamenti del mondo reale [2]. Rilevare volti, almeno frontalmente, non è più un compito difficile; per questo motivo, la ricerca in questo campo si sta focalizzando sul miglioramento del riconoscimento facciale in condizioni stimolanti [3].

In questo lavoro di tesi è stato creato un dataset partendo da un banco di fotosegnalamento. Il banco in questione è formato da un braccio meccanico di posizionamento al di sopra del quale sono state fissate quattro telecamere. In questo modo si riesce ad ottenere un sistema di fotosegnalamento in grado di attuare foto segnaletiche da diversi punti di vista. Inoltre, è stata creata una struttura con altre cinque telecamere, poste ad altezze e inclinazioni diverse, in modo da simulare un ambiente in cui è presente un sistema di videosorveglianza nel miglior modo possibile. Per il controllo e la movimentazione delle telecamere, nel caso del fotosegnalamento, è stato utilizzato *LabView*. Tramite questo software è stato possibile ottenere come output per ogni soggetto 28 foto segnaletiche, da diverse prospettive. Per l'acquisizione dei video, invece, è stato utilizzato *OBS studio*, che ha permesso la ripresa del soggetto dalle cinque telecamere, ottenendo un video per ogni telecamera e un filmato complessivo che mostra l'intera azione.

Scopo della Tesi

Lo scopo della tesi è quello di realizzare un dataset diverso rispetto a quelli già presenti in letteratura.

Il dataset ottenuto, infatti, presenta foto segnaletiche dei soggetti da numerosi punti di vista, abbinate a dei video che mirano all'accurata riproduzione di un ambiente con videosorveglianza. Inoltre, alcune persone sono state riprese con oggetti che fornivano una diversa copertura del volto, in modo sia da simulare il quotidiano sia per mettere alla prova il software di riconoscimento facciale e la sua accuratezza.

Struttura della Tesi

La tesi presenta diversi capitoli all'interno dei quali vengono spiegate le procedure e i materiali utilizzati sia nella parte di fotosegnalamento che nella simulazione di un ambiente con videosorveglianza:

- 1- Il primo capitolo include lo stato dell'arte: sono stati illustrati i diversi dataset e ambienti di videosorveglianza (elencando le videocamere maggiormente utilizzate) presenti in letteratura ed evidenziate le differenze con quello ottenuto da questo lavoro.
- 2- Nel secondo capitolo sono stati descritti i materiali e i metodi utilizzati per la creazione del dataset, spiegando in dettaglio i software e le loro funzionalità.
- 3- Il terzo capitolo racchiude i diversi protocolli utilizzati durante l'acquisizione di un soggetto. Sono stati divisi in: protocollo per l'acquisizione delle foto, protocollo video e, in più, è stato interamente riportato il modulo di dichiarazione di consenso della privacy compilato da ogni soggetto sottoposto allo studio.
- 4- Nel quarto e ultimo capitolo, è stata attuata un'analisi dei risultati tramite una tabella, la quale racchiude tutte le acquisizioni attuate per ogni soggetto presente nel dataset.

Capitolo 1 Stato dell'arte

In questo capitolo è stato illustrato lo stato dell'arte attuale sul fotosegnalamento e sui meccanismi utilizzati per la videosorveglianza. Oltre alla descrizione di ciò che viene maggiormente usato, è stata attuato anche un confronto con il caso studio, in modo da evidenziarne i punti di forza e di debolezza.

1.1 Fotosegnalamento

Negli ultimi anni, l'intelligenza artificiale ha fatto grandi miglioramenti in molti settori. Tra questi, risalta l'utilizzo da parte delle forze dell'ordine, le quali hanno sfruttato l'avanzamento delle tecnologie per indagare su attività sospette. Un esempio è quello del controllo degli incidenti stradali: utilizzando le telecamere di sorveglianza, infatti, entra in gioco l'utilità del riconoscimento facciale sfruttando i dataset di fotosegnalamento [7]. Nell'ultimo periodo il riconoscimento facciale è stato largamente utilizzato in diversi settori, ad esempio, nel campo degli smartphone, l'autenticazione biometrica è andata a sostituire l'impronta digitale per autorizzare diverse attività, come ad esempio pagamenti o il semplice sblocco dell'apparecchiatura.

In settori come quello della Polizia di Stato, in cui a volte era disponibile solamente l'immagine del ricercato, ma non la sua impronta digitale, è stato fondamentale per il miglioramento del riconoscimento la creazione di dataset come il *S.A.R.I.* (Sistema Automatico Riconoscimento Immagini). Quest'ultimo, partendo da un'immagine fotografica di un soggetto ignoto, riesce a effettuare una ricerca computerizzata nella banca dati *A.F.I.S.* (Automatic Fingerprint Identification System) e grazie a due algoritmi di riconoscimento facciale, è in grado di fornire un elenco di immagini dalla più simile alla meno. La problematica principale è che il fotosegnalamento, attuato dal corpo di Polizia, consiste nello scattare solamente tre foto (una frontale e due di profilo) che, in molti casi, non permettono di trovare un match in immagini con minore risoluzione che possono essere state catturate, ad esempio, da telecamere di sorveglianza. L'attrezzatura da loro utilizzata è quella prodotta dalla *Secom S.r.l.*, che permette, in un unico scatto, l'acquisizione sia della foto frontale che di quelle di profilo. Proprio in questo ambito subentra il tipo di fotosegnalamento trattato in questa tesi: la differenza principale rispetto al fotosegnalamento classico descritto in precedenza, infatti, riguarda il numero di foto acquisite e le diverse angolazioni.

Il meccanismo di acquisizione è stato creato a partire da un braccio meccanico di posizionamento, al di sopra del quale sono state poste 4 *Raspberry Pi Zero W* in diverse inclinazioni e posizioni. Questo meccanismo è in grado di acquisire quattro fotogrammi in sette posizioni diverse, in modo da scansionare l'intero volto del soggetto sottoposto alla procedura. Sulla base di questo è stato creato un nuovo dataset che include 46 soggetti con foto segnaletiche scattate da 28 prospettive diverse, insieme a cinque video (utilizzati per simulare le riprese di videosorveglianza). In questo modo, il maggior numero di acquisizioni del volto del soggetto in diverse angolazioni permette di aumentare la possibilità di riscontro nel riconoscimento facciale tra la foto segnaletica e la videosorveglianza, andando a migliorare il sistema attualmente in utilizzo.

Per evidenziare le differenze con i database presenti in letteratura, si può esaminare l'*ORL* (Olivetti Research Laboratory). Esso prende in considerazione 40 soggetti, fornendo solamente una foto frontale, ed ognuno di questi caratterizzato da una diversa illuminazione ed espressioni facciali. L'unica caratteristica in comune in tutte le acquisizioni è la presenza di uno sfondo nero, ma la variazione delle condizioni non consente l'ottenimento di risultati confrontabili [8]. Allo stesso modo per *YouTube Faces Database*, è stato preso in considerazione un ampio gruppo di soggetti, dato che le immagini sono state acquisite dal web stesso. L'obiettivo era quello di prendere una coppia di immagini e cercare se ci fosse un match o meno tra le due. Anche in questo caso, esso non risulta adeguato al riconoscimento facciale partendo da fotogrammi di videosorveglianza, perché non si è sempre in grado di distinguere le caratteristiche biometriche anche a causa della presenza di uno sfondo non neutro [9].

Per concludere, nessuno dei due database include set di foto segnaletiche e video di sicurezza comparabili, per questo non possono essere utilizzati per valutare l'impatto dell'utilizzo di foto segnaletiche da diversi punti di vista in scenari di videosorveglianza.

Alcune banche dati che includono soggetti con pose differenti e foto segnaletiche da diverse prospettive, sono presenti in letteratura. Per esempio, il *Facial Recognition Technology* (FERET) ha all'interno 14051 immagini a colori (di dimensioni 512x768 pixels) di 1199169 soggetti. Per 200 soggetti sono presenti nove foto segnaletiche acquisite da differenti punti di vista (da -60° a $+60^\circ$). Anche in questo caso non sono presenti video abbinati in grado di simulare un'area di videosorveglianza, inoltre il database non è accessibile a chiunque [4]. Dalla Figura 1.1 si nota come nel *FERET* non sempre il soggetto mantenga la stessa mimica facciale durante l'acquisizione e come le foto non abbiano un ordinamento a priori.



Figura 1.1 - Esempio di un soggetto presente all'interno del database FERET [11].

1.2 Videosorveglianza

La necessità di assicurare il bene pubblico è sempre in aumento, di conseguenza diversi progetti, sia industriali che accademici, si sono iniziati a focalizzare nella creazione di sistemi per monitorare gli spazi pubblici.

Molte proposte comprendevano l'utilizzo di telecamere wireless, le quali sono diventate più disponibili, accessibili ed affidabili, e riescono a bilanciare maggiormente la potenza con l'efficienza energetica.

I *WVSN* (Wireless Visual Sensor Network) presentano diverse applicazioni di sorveglianza in scenari differenti (per monitorare l'habitat oppure come risposta di emergenza o videosorveglianza utilizzata dalle forze dell'ordine).

Ci sono diversi requisiti da rispettare sia per telecamere wireless interne che esterne:

- efficienza energetica
- distribuzione dei dispositivi in zone con banda internet densa
- hardware efficiente
- meccanismi che assicurano la privacy dell'individuo.

Il problema principale è che registrare continuamente, soprattutto per quanto riguarda la videosorveglianza esterna, è molto dispendioso [6]. Per questo vengono aggiunti sensori che permettono l'identificazione di oggetti specifici o che sono sensibili al movimento [10].

Un'altra tecnica utile per incrementare il risparmio energetico è quello di utilizzare una risoluzione più bassa per l'acquisizione video, come riportato nella prima colonna della Figura 1.2.

Table 1. Taxonomy of current wireless visual sensor networks (WVSNs) for public-space surveillance.								
System	Energy efficiency	Multimodal	Bandwidth efficiency	Wireless	Computer vision	System software	Camera overlap	Security
Citric power source: battery powered; frame rate and resolution: 15 fps at 1,280 × 1,024; 30 fps at 640 × 480	Sleeping nodes wake on audio; efficient image sensors and radios	Audio	In-network image processing; small image size	802.15.4	Compressed sensing; frame differencing; background subtraction; multiple cameras per object	Embedded Linux	Yes	Data encrypted at node, com- munication protocol
HuSIMS power source: grid power; frame rate and resolution: fps at 320 × 240	Sleeping nodes; small processing power cameras; power-efficient sensors	Smoke, motion, light, humidity, and temperature detectors	In-network image process- ing; sink nodes; separate low- bandwidth sensor networks	WiMAX 4G	Alerts for abnor- mal object behavior	Large man- agement software with semantic processing	No	Communica- tion protocol
OmniEye power source: battery power; frame rate and resolution: 30 fps at 640 × 480	Same as Citric and efficient MAC layer; separate energy- efficient sensor network	Tempera- ture, light, sound	In-network image processing; data annotation	802.11	JPEG compres- sion (70%); adaptive back- ground subtraction	µLinux	Yes	Data encrypted at node; com- munication protocol

Figura 1.2 - Sensori maggiormente utilizzati per la videosorveglianza e loro principali caratteristiche.

A partire dalla Figura 1.2 è possibile notare come il sistema operativo più utilizzato per sorveglianza wireless sia *Linux*, grazie al suo hardware indipendente. Invece il device più popolare è *Raspberry Pi*, che permette di avere una capacità di comunicazione elevata, in particolare con *Linux*.

Un'applicazione largamente utilizzata in questo campo è *SWEETcam*: questa utilizza *Raspberry Pi* che permette di catturare 1920x1080 immagini a colori a 30 fps. I sensori di immagine possono essere anche collegati tramite un USB che riduce istantaneamente il consumo di energia da parte della telecamera stessa.

Per quanto riguarda, invece, la videosorveglianza all'interno di filiali bancarie, molte di queste, per tutelare gli utenti da possibili furti o aggressioni, posizionano i terminali ATM in aree controllate da telecamere a circuito chiuso, l'accesso alle quali è regolato da porte automatiche.

L'apertura delle porte deve essere anch'essa controllata e permessa solamente a chi deve effettuare operazioni agli sportelli automatici, cioè agli effettivi utenti del sistema.

L'accesso all'area controllata deve perciò avvenire in modo selettivo, permettendo di accedere agli utenti in possesso di carte di credito o debito di vari circuiti.

In Figura 1.3 si ha un esempio della visualizzazione tramite una microcamera *IPC0122P3*, più comunemente chiamata *pin hole*. È una telecamera nascosta ad alta definizione con risoluzione 1920x1080 (Full HD) a 30 fps (real time). La *pin hole* è stata sviluppata per essere installata nei bancomat o ATM, ossia dove c'è bisogno di una telecamera nascosta e di ottima risoluzione per permettere la visione dettagliata della persona e riconoscere eventuali truffe.



Figura 1.3 – Immagine di una telecamera posta all'interno di un bancomat.



Figura 1.4 – Immagine estrapolata dal video acquisito dalla telecamera 4 in laboratorio.

Invece, per quanto riguarda la telecamera utilizzata nella riproduzione di un ambiente di videosorveglianza in laboratorio, si tratta di una webcam *Irarucw* con una risoluzione di 1080p, capace di acquisire immagini ad una velocità di 30 fps, posizionata in modo da simulare quella utilizzata nei bancomat.

Si può notare come le riprese abbiano una diversa inclinazione: nella Figura 1.3 si ha una visione dall'alto che arriva quasi ad oscurare il volto, mentre nella Figura 1.4 si ha una visualizzazione del soggetto ad un'altezza minore con il profilo maggiormente visibile.

Capitolo 2 Materiali e metodi

All'interno del capitolo vengono descritti i materiali e i metodi utilizzati nello studio illustrato in questa tesi. Innanzitutto, c'è una descrizione generale del banco, con un'illustrazione degli elementi fondamentali usati nella sperimentazione. Inoltre, è fornito un elenco dettagliato delle caratteristiche degli strumenti insieme alle loro modalità di uso. Infine, vi è una trattazione dei due software fondamentali: uno utilizzato nel fotosegnalamento e l'altro per la simulazione di un ambiente fornito di videosorveglianza.

2.1 Descrizione del banco

Per la creazione del dataset discusso in questa tesi è stato utilizzato un banco composto da:

- una cabina all'interno della quale è posta una seduta dove far posizionare il soggetto da sottoporre al fotosegnalamento
- un braccio di posizionamento rotante provvisto di telecamere per acquisire le immagini da varie angolazioni
- una struttura dotata di telecamere che permette di simulare la videosorveglianza in diverse accezioni (come, ad esempio, al di sopra delle vetrine dei negozi oppure all'interno degli ATM).

Sono state acquisite foto segnaletiche a partire da sette angoli nel piano orizzontale e quattro in quello verticale. In dettaglio, nel piano orizzontale, le foto sono state prese da angolazioni da -135° a $+135^\circ$ con uno step di 45° (dove a 0° la telecamera è posta sul piano degli occhi del soggetto – Figura 2.3). Nel piano verticale, invece, le acquisizioni sono state attuate da -30° a $+60^\circ$, con 30° di step. Lo sfondo di ogni foto segnaletica è rappresentato da un fondo nero e le quattro telecamere sono state poste su un braccio robotico che ruota attorno all'asse verticale. Questa rotazione ha permesso l'acquisizione delle foto segnaletiche dai sette assi sul piano orizzontale e quattro sul piano verticale [7]. È stato così possibile scattare, nello stesso istante, foto a gruppi di quattro, dove, per ogni gruppo, le acquisizioni presentavano la medesima posizione di rotazione del braccio.

Per quanto riguarda l'illuminazione, il braccio rotante è stato dotato di una striscia di led omogenea su tutta la superficie frontale al soggetto. Invece, nel caso della riproduzione di ambienti di videosorveglianza, la struttura creata è stata posta in modo da sostenere cinque videocamere ad angolazioni ed altezze diverse. La procedura è stata creata in modo da

riprendere il soggetto sia frontalmente che di profilo, per semplificare il successivo riconoscimento facciale.

Per attuare il protocollo è stato utilizzato *LabView* come software di controllo, movimentazione e acquisizione telecamere e *OBS studio* per l'acquisizione dei video.

2.2 Descrizione telecamere

Per la realizzazione del fotosegnalamento e per la simulazione di un ambiente dotato di videosorveglianza sono stati utilizzati tre tipi diversi di telecamere, ognuna con le sue specifiche:

- *RASPBERRY PI ZERO W*

Questa tipologia di telecamera è stata utilizzata come strumento di cattura principale di foto utili per il fotosegnalamento. La scelta effettuata non riguarda solo le caratteristiche tecniche, ma presenta ulteriori vantaggi: la semplicità del montaggio, la facilità di messa a fuoco e la leggerezza. L'ultima specifica citata è fondamentale perché le telecamere in questione sono state montate su un supporto tubolare, facente parte del braccio di posizionamento; grazie a questa caratteristica riescono a non appesantirne la meccanica.

Più nello specifico, *Raspberry Pi Zero W* è un piccolo device che può essere utilizzato in vari modi, tra i quali come telecamera. È equipaggiato da un chip wireless chiamato *Cypress CYW43438* (Figura 2.1) che permette di avere 802.11 n wireless LAN e connessione Bluetooth 4.0.

Le specifiche sono le seguenti:

- dimensioni: 65x 30x5 mm
- SoC: Broadcom BCM 2835 chip
 - ARM11 a 1 GHz, single-core CPU
 - 512 MB RAM
- storage: MicroSD card
- video e audio: 1080 p HD video e stereo audio via mini connettore HDMI
- potenza: 5V tramite connessione con micro-USB
- wireless: 2.4 GHz 802.11 n wireless LAN
- Bluetooth: Bluetooth classic 4.0 e Bluetooth Low Energy (BLE)
- output: Micro-USB
- GPIO: 40-pin GPIO, unpopulated



Figura 2.1 - Chip della *Raspberry Pi Zero W* utilizzata per l'acquisizione di immagini nel fotosegnalamento.

La costruzione di un dataset fornito di foto segnaletiche necessita l'utilizzo una telecamera, in questo caso è stata montata a lato del chip utilizzando un cavo. I device possono essere utilizzati tramite rete wireless, infatti è stato possibile scaricare le foto acquisite dal web e salvarle su un computer sfruttando il software di controllo movimentazione e acquisizione telecamere (presente su *LabView*).

Per proteggere i componenti, è utile usufruire della custodia protettiva in dotazione (Figura 2.2), la quale espone solo le parti necessarie.



Figura 2.2 - Custodia di protezione della *Raspberry Pi Zero W* utilizzata per proteggere i componenti interni.

Per ogni acquisizione di un soggetto nel fotosegnalamento sono state collezionate foto segnaletiche da sette angoli nel piano orizzontale e quattro nel piano verticale. Quest'ultimo piano è rappresentato dalle *Raspberry Pi Zero W* che sono state poste sul braccio di posizionamento a partire da un angolo di -30° fino ad un angolo di $+60^\circ$ con 30° per ogni step, dove 0° è frontalmente al soggetto (Figura 2.3).

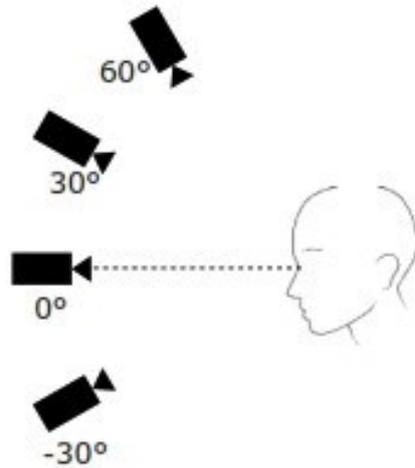


Figura 2.3 - Punti di vista sul piano verticale utilizzati per ottenere le foto segnaletiche.

- *WEBCAM AIGOSS*

Questa telecamera aggiuntiva è stata introdotta all'interno della cabina dove viene attuato il fotosegnalamento per garantire un'acquisizione del viso del soggetto perfettamente frontale. È stata posizionata in modo che potesse comprendere nella totalità il volto della persona fotosegnalata ed è stata successivamente fissata. Proprio per le sue caratteristiche costruttive garantisce un'acquisizione con una risoluzione di 1080 p.

Questa telecamera è stata utilizzata ulteriormente per verificare che il soggetto fosse posto alla perfezione all'interno di un riquadro fittizio, in modo da comprendere l'intero volto nell'acquisizione successiva. Infine, essa è stata anche considerata un punto di riferimento: durante la procedura è stata indicata come punto da guardare, in modo che il soggetto non venisse distratto dal braccio di posizionamento in movimento, oppure infastidito dalle luci led poste sul tubolare.

Per garantire una corretta illuminazione del viso del soggetto durante la foto frontale attuata da questa telecamera, è stata inserita una lampadina led a luce calda all'interno della cabina. Questa veniva comandata dalle operatrici dall'esterno, ed accesa solo nel momento di acquisizione della foto frontale stessa.

- *WEBCAM IRARUCW*

Per la simulazione di un ambiente dotato di videosorveglianza, sono state utilizzate cinque telecamere poste in diverse angolazioni in modo da riprodurre le posizioni delle diverse videocamere di sicurezza (in modo da simulare quelle poste all'interno dei Bancomat oppure utilizzate al di fuori dei negozi). La webcam in questione ha anch'essa una risoluzione di 1080 p, capace di acquisire immagini ad una velocità di 30 fps. Nella procedura di acquisizione video, la risoluzione è stata personalizzata a 352x288 tramite *OBS studio*, in modo da garantire una continuità nei cinque video diversi e per evitare che occupassero troppa memoria all'interno del computer (in media 10 MB per video). Queste telecamere sono dotate di un cavo usb per agevolare il collegamento con un computer. A causa del loro numero è stato necessario l'utilizzo di un hub per aumentare il numero di porte d'accesso nel computer utilizzato, ma questo accorgimento non ha influito in alcun modo con l'acquisizione.

È stata costruita una struttura apposita che prevedeva l'inserzione delle cinque telecamere con orientazioni e altezze differenti. La struttura è stata posizionata all'interno del laboratorio dove sono stati creati dei segni sul pavimento così da riposizionarla al meglio nel caso fosse stata spostata, in modo da evitare una possibile non riproducibilità tra le foto dei vari soggetti. La forma scelta della struttura è stata quella di un arco quadrato, sia per la facilità di costruzione, sia per non ostacolare il protocollo video.

E' stata scelta questa forma specifica in modo da permettere la riproduzione il più fedele possibile ai diversi tipi di videosorveglianza, sia interna che esterna agli edifici. Per esempio, alcune telecamere sono state poste ad altezze non indifferenti, circa 2.10 m, in modo da simulare la videosorveglianza stradale oppure quella all'interno di negozi, mentre altre hanno una visione più frontale circa 1.50 m (Figura 2.4) ed infine l'ultima è posta ad un'altezza ancora minore, in modo da essere fedele con le telecamere presenti all'interno degli ATM.



Figura 2.4 - Struttura utilizzata per la riproduzione della videosorveglianza in laboratorio.

2.3 Descrizione braccio di posizionamento (meccanica e movimentazione)

Il braccio di posizionamento è stato creato a partire da un telaio di una bicicletta sul quale è stato installato un tubolare. È dotato di un motore passo passo, ossia un motore elettrico sincrono che mantiene con precisione la propria velocità di rotazione e posizione. Sono motori che, a differenza di tutti gli altri, hanno come scopo quello di mantenere fermo l'albero in una posizione di equilibrio. Questa caratteristica è fondamentale per avere una stabilità del prodotto finale e il braccio in una ben precisa posizione angolare. Nel caso studio riportato in questa tesi, non è stato possibile imprimere un movimento di un arco minore rispetto al percorso totale, e poi stoppare il meccanismo, dato che la meccanica del braccio lo impediva. Infatti, quest'ultimo presenta una struttura troppo pesante ed è stato notato che, nel compiere uno spostamento piccolo e poi fermarsi, le oscillazioni date dal freno erano maggiori dell'angolo in input e quindi non permettevano di continuare la procedura. Per questo motivo, è stato scelto un movimento continuo del braccio meccanico, con telecamere programmate a compiere acquisizioni durante il raggiungimento di specifiche posizioni.

Le foto sono scattate da quattro telecamere posizionate sul tubolare in quattro posizioni diverse. Il braccio di posizionamento ruota attorno un asse verticale e compie sette angoli

nel piano orizzontale che corrispondono alle posizioni di acquisizione. Una schematizzazione del braccio robotico è rappresentata nella Figura 2.5.

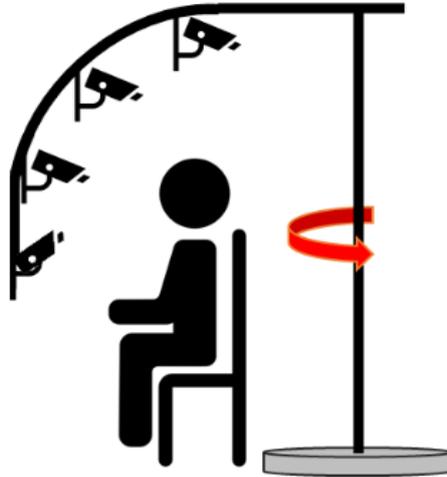


Figura 2.5 - Funzionamento schematizzato del braccio di posizionamento.

È stato fissato un punto di inizio corsa in modo da far corrispondere la foto centrale con la metà del percorso del braccio robotico, e per evitare che quest'ultimo finisse erroneamente contro la parete posteriore della cabina. Sono stati fatti svariati tentativi alla fine dei quali, guardando l'output (in particolare, la foto 33 - foto frontale), a seconda di quale parte di profilo era stata acquisita, veniva spostato il punto di inizio corsa. Più nello specifico, se nella foto fosse prevalso il profilo destro del soggetto, il braccio avrebbe dovuto essere spostato in avanti; viceversa se fosse prevalso il profilo sinistro. Una volta ottenuta la foto 33 perfettamente frontale, all'interno della cabina è stato inserito un pilastro in modo da avere un punto di riferimento e rendere le acquisizioni riproducibili.

Per consentire un'illuminazione ottimale del soggetto, sulla superficie del braccio di posizionamento, sono state inserite delle stringhe luminose a LED con temperatura di colore di circa 3500 K, schermate con carta bianca da 80 g/m² per diffondere la radiazione, evitando così la sovraesposizione.

Per posizionare al meglio il soggetto all'interno della cabina è stato inizialmente visualizzato, tramite il collegamento con le telecamere, se il volto fosse all'interno di tutte le inquadrature. Una volta trovata l'altezza ottimale, è stato inserito un rivestimento in gommapiuma sulla parete posteriore della cabina. Oltre che come punto di riferimento, è stato utilizzato per garantire una stabilità del capo del soggetto ed evitare che ci fossero movimenti involontari che potessero causare delle acquisizioni imprecise.

2.4 Descrizione software di controllo movimentazione e acquisizione telecamere (LabView)

LabView è un ambiente di programmazione grafica utilizzato per sviluppare sistemi automatizzati di ricerca e test di produzione. Introdotto nel 1986, fa parte del gruppo NI (National Instruments) e il suo scopo principale è quello di creare un software per controllare gli strumenti ad esso collegato, automatizzare i test e i sistemi di misurazione [5]. Il linguaggio utilizzato è stato definito *graphical, structured data-flow language*:

- *graphical*, dato che è rappresentato come un diagramma bidimensionale di nodi e fili interconnessi al posto del testo
- *structured* si riferisce all'uso delle boxe per rappresentare i vari controlli, i loop e le esecuzioni condizionali. Infatti, all'interno di queste strutture sono presenti dei diagrammi che controllano se e quando eseguire un comando.
- *data-flow* indica la semantica del linguaggio.

Il programma utilizzato per il fotosegnalamento, presenta un *front panel* dove si visualizzano gli input principali e i rispettivi output, e un *block diagram* ossia l'ambiente di programmazione.

Nella descrizione del *front panel* è possibile dividere i diversi input (Figura 2.6):

- 1- numero di angoli: corrisponde al numero delle posizioni che il braccio di posizionamento deve assumere per poi acquisire l'immagine (nel caso in questione sono sette)
- 2- tempo tra ogni scatto; è stato impostato a 1000 ms in modo da permettere che ci sia una stabilizzazione del meccanismo per ottenere delle immagini a fuoco
- 3- *serial port*, che collega il meccanismo del braccio di posizionamento al computer utilizzato e quindi è alla base del funzionamento del fotosegnalamento in questione. È una caratteristica che cambia da computer a computer, quindi, bisogna controllarla in modo da garantire il collegamento stesso

- 4- *path*: rappresenta il percorso di salvataggio delle immagini su una cartella a scelta dall'operatore. Questo accorgimento può essere utilizzato per distinguere con facilità i soggetti che si sottopongono alla procedura.

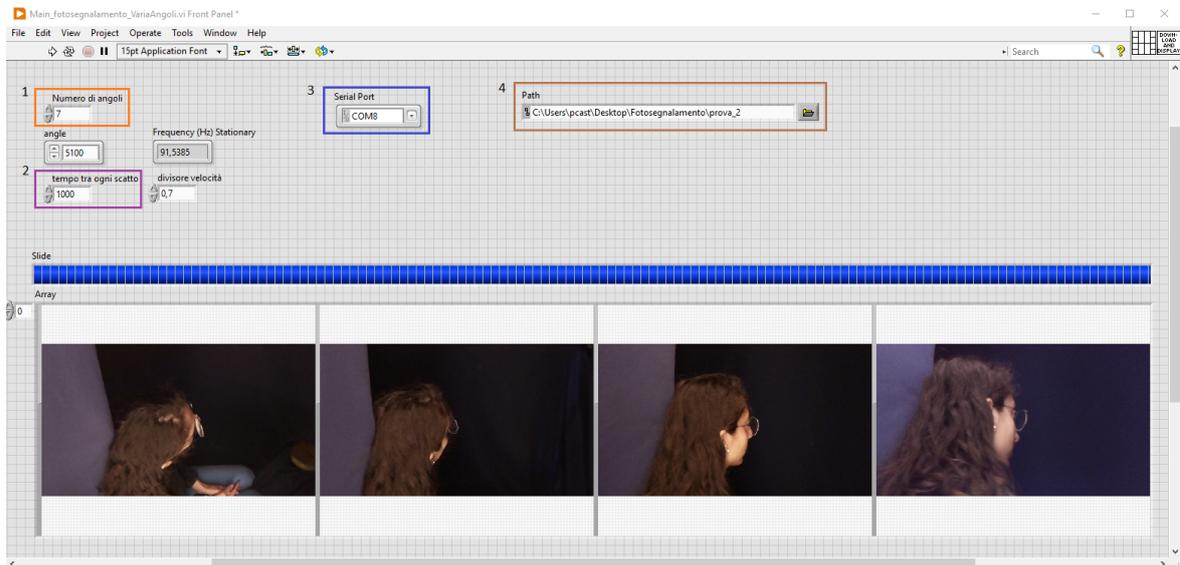


Figura 2.6 - Immagine del front panel con evidenziati gli input.

Sono visionabili anche due output (Figura 2.7):

- 1- slide: un output fittizio rappresentato da una barra che permette di visionare, tramite una successione di barrette blu, il tempo restante per terminare il processo. Esso è definito fittizio perché permette semplicemente di avere un'idea e fornisce un feedback del funzionamento della procedura di fotosegnalameto
- 2- foto segnaletiche, l'output vero e proprio, scattate dalle 4 *Raspberry Pi Zero W* nei diversi istanti di tempo. Queste sono visionabili a gruppi di quattro direttamente sul *front panel* per ogni posizione acquisita, e poi salvate sul *path*.

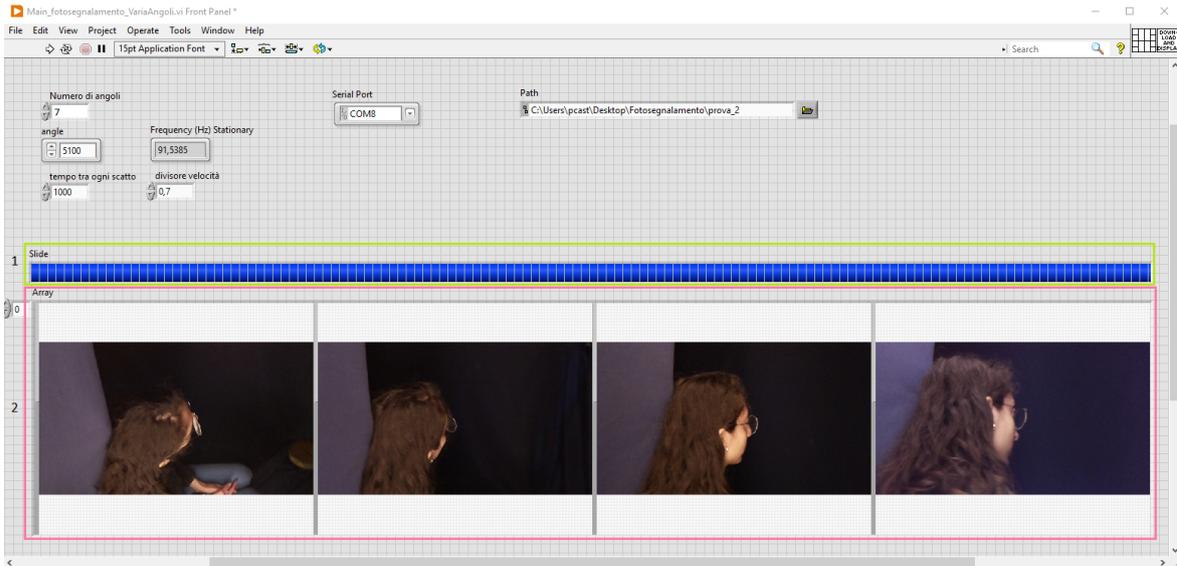


Figura 2.7 - Immagine del front panel con evidenziati gli output.

Il *block diagram* è diviso in due ambienti principali descrivibili separatamente:

- 1- nel primo (Figura 2.8) è presente un ciclo *while* che prende come input quelli indicati in Figura 2.7 e, creando una rampa, permette il movimento del braccio di posizionamento. Principalmente è stato programmato in modo da selezionare un istante di tempo in cui il meccanismo rimane fermo per scattare le foto per poi partire nuovamente e raggiungere la posizione successiva. Appena quest'ultimo raggiunge lo step finale, è richiamato indietro per tornare nella posizione iniziale pronto per una nuova acquisizione.
- 2- l'altro ambiente principale (Figura 2.9) è provvisto di un ciclo *for* che, avendo come input principale il numero di angoli, permette il salvataggio delle foto sul *path* per gli n angoli scelti (nel caso in questione sette). All'interno di questo blocco sono stati posti gli indirizzi di ogni *Raspberry Pi Zero W* e, ognuno di questi, è legato a *download and display*. Il *tool* in questione è in grado di scaricare l'*array* di foto dal web e mostrarle come output sul *front panel*. Successivamente viene sistemata la risoluzione verticale e orizzontale dell'immagine prima di essere inserita all'interno di *concatenate strings*. Quest'ultimo *tool* permette di unire le quattro foto acquisite da ogni telecamera in una specifica posizione, scegliere in che formato viene salvato (in questo caso è stato scelto JPG) e nominarle da 1 a 4 per riconoscere a che telecamera si fa riferimento. Infine, viene utilizzato il

comando *write file* in grado di scrivere l'immagine su un file e salvarla all'interno del computer nel formato selezionato (nel nostro caso sempre JPG).

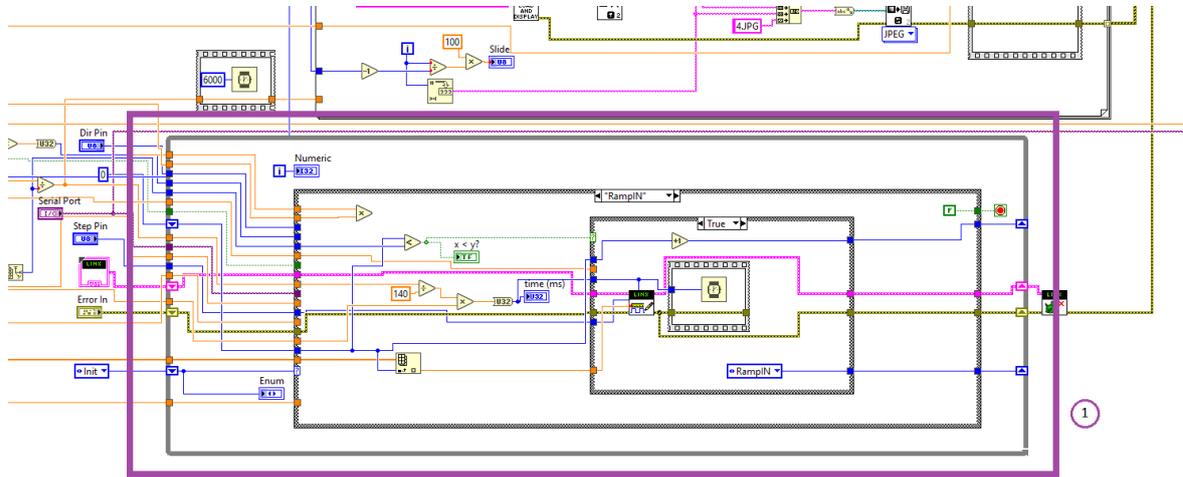


Figura 2.8 - Immagine del block diagram che mostra il ciclo while per il controllo del motore che mette in rotazione il supporto delle telecamere.

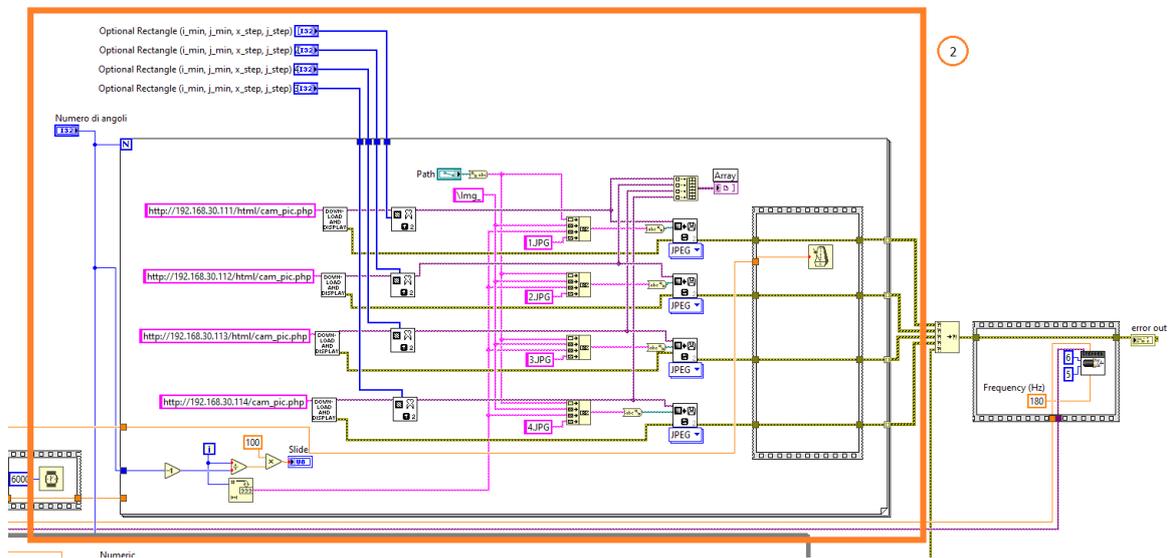


Figura 2.9 - Immagine del block diagram che mostra il ciclo for per il controllo delle telecamere.

2.5 Descrizione del software di acquisizione video (OBS studio)

OBS Studio è un software per registrare video e per attuare live streaming. Per quanto riguarda l'utilizzo dell'audio, come default il programma cattura l'audio del desktop e il microfono del computer utilizzato (in alternativa esso può essere settato manualmente). Nel caso del video, invece, all'apertura del programma si presenta lo schermo interamente nero, questo è dato dal fatto che il software necessita di una sorgente per catturare video.

Prima di utilizzare *OBS Studio* bisogna configurarlo, per quanto riguarda il caso in studio sono stati attuati diversi passaggi:

- 1- configurazione webcam *Irarucw*: bisogna cliccare sul box denominato “*sources*”, cliccare sul “+” e scegliere il dispositivo che si vuole utilizzare tra le periferiche collegate al computer (Figura 2.10). In questo caso specifico, le telecamere predisposte sono cinque, numerate fisicamente sul dispositivo stesso (tramite un adesivo) e, allo stesso modo, nel software, ad esempio la camera 1, è indicata come *Cam1*.

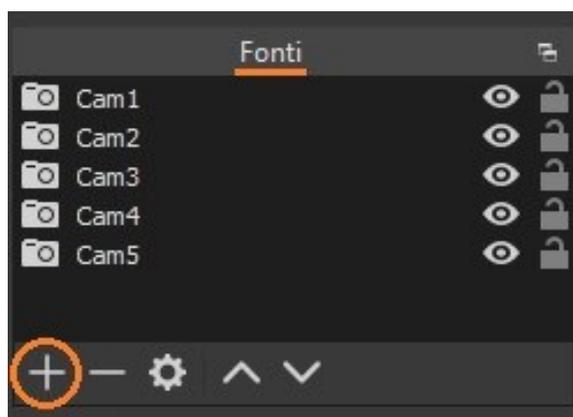


Figura 2.10 – Elenco delle sorgenti per la cattura del video.

- 2- modifica proprietà: visualizzando le “proprietà” per ogni telecamera, è stata modificata la risoluzione a 352x288 (Figura 2.11).

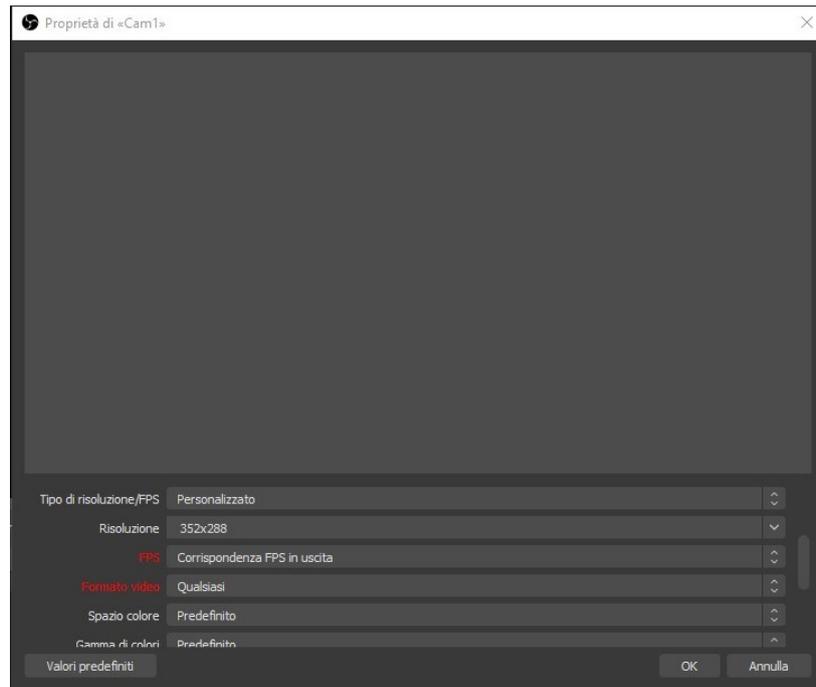


Figura 2.11 - Proprietà di una telecamera all'interno di OBS studio.

3- ordinamento telecamere: per migliorare la visualizzazione si è pensato fosse utile ordinarle a partire da in alto a sinistra fino in basso a destra. In questo modo i cinque riquadri corrispondono ai video catturati rispettivamente dalla telecamera 1 fino ad arrivare alla telecamera 5 (Figura 2.12). Lo scopo è quello di ottenere un video che comprenda la visione da tutte e cinque le telecamere ed altri cinque video che ritraggono ognuno la visione da una videocamera differente.



Figura 2.12 – Ordinamento delle telecamere.

- 4- *source record*: questo *tool* aggiuntivo, scaricabile dal web, permette di avere sei video in totale. Grazie a questo è possibile avere come risultato finale i cinque video separati oltre che il video che ritrae l'azione da tutte le telecamere contemporaneamente. Per inserirlo bisogna configurare ogni telecamera, andare su “filtri” e, nella sezione “filtri degli effetti”, selezionare “*source record*” (Figura 2.13).

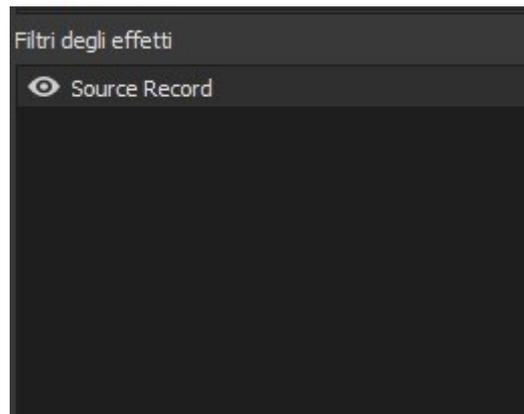


Figura 2.13 - Visualizzazione finale dopo l'applicazione del filtro “*source record*”.

- 5- *selezione cartella*: per gestire la cartella nella quale verranno salvati i video, bisogna andare su “impostazioni”, selezionare “uscita” poi “percorso di registrazione” e, tramite sfoglia, inserire la cartella prescelta (Figura 2.14).

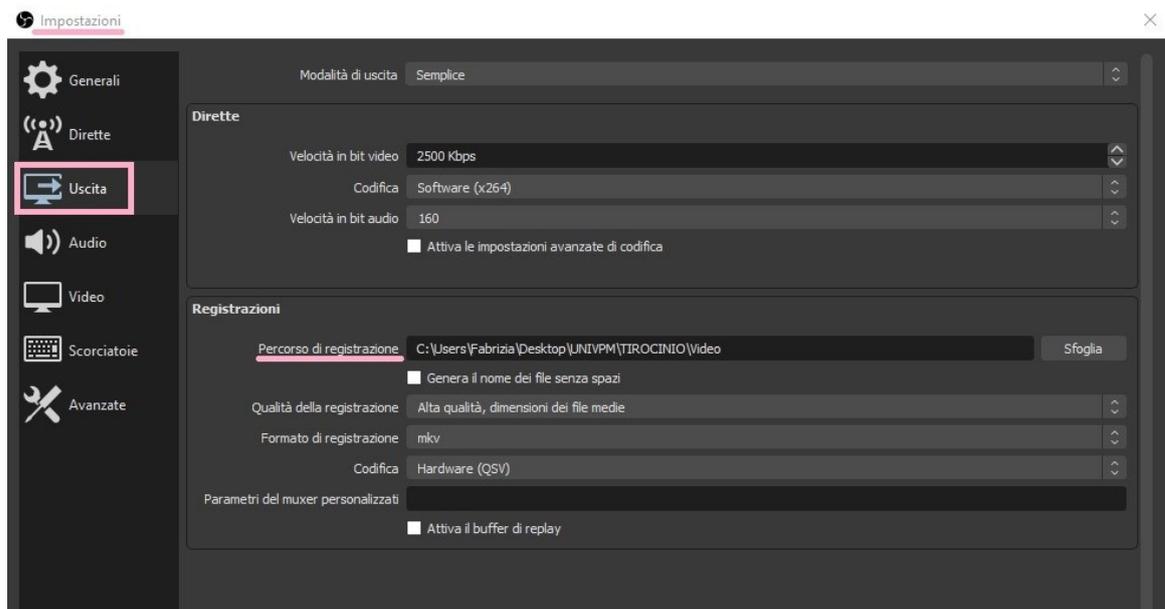


Figura 2.14 – Struttura dei passaggi per selezionare la cartella dove salvare i video.

- 6- acquisizione schermo: dopo aver settato correttamente il programma, per iniziare un'acquisizione bisogna premere su “avvia la registrazione”, nella sezione “controlli”, presente sulla schermata iniziale del software in basso a destra (Figura 2.15).



Figura 2.15 - Posizione del comando utile ad avviare la registrazione.

Capitolo 3 Sperimentazione

In questo capitolo vengono fornite maggiori indicazioni riguardo i protocolli utilizzati per l'acquisizione di foto tramite braccio di posizionamento dotato di quattro *Raspberry Pi Zero W*, in differenti posizioni e angolazioni, e di video tramite camere distribuite in modo da ricoprire lo spazio indicato simulando telecamere di videosorveglianza. In aggiunta, è indicato il protocollo di dichiarazione di consenso alla privacy, necessario per effettuare qualsiasi tipo di studio che riguardi l'utilizzo di immagini di una persona.

3.1 Protocollo acquisizione foto

- 1- il soggetto viene fatto entrare all'interno della cabina (con nessun elemento sul viso che potrebbe interferire con la misurazione) e si posiziona sulla seduta
- 2- si controlla che la testa sia poggiata sul supporto in gommapiuma e che sia posizionato frontalmente rispetto alla telecamera interna
- 3- il braccio meccanico viene portato a inizio corsa (corrispondente al pilastro metallico posto all'interno della cabina)
- 4- la cabina viene opportunamente chiusa in modo da oscurare la luce esterna
- 5- dalla postazione computer l'operatore accende la luce interna per controllare che il soggetto sia posto in posizione frontale e che si vedano entrambe le porzioni di viso in modo simmetrico
- 6- per ogni soggetto si crea una cartella "Nome_Cognome" del fotosegnalato, al cui interno vengono poste: cartella "Nome_Cognome_Fotosegnalamento" contenente le immagini per l'acquisizione 3D del volto, cartella "Nome_Cognome_Video" per la simulazione delle telecamere di sicurezza poste a diverse altezze, "Nome_Cognome.jpeg" in riferimento all'immagine frontale realizzata con la fotocamera interna e un file testo "Nome_Cognome.txt" con dati anagrafici riguardanti il sesso e l'età
- 7- tramite la telecamera interna si scatta la foto frontale con la luce accesa
- 8- la luce viene spenta e viene avviato il braccio meccanico in modo da iniziare l'acquisizione
- 9- il braccio meccanico è costituito da quattro *Raspberry Pi Zero W* orientate secondo diverse angolazioni in modo da scansionare il volto nella sua interezza. Nel percorso, che parte da un palo indicante il fine corsa, acquisisce immagini da sette angolazioni

con un tempo tra uno scatto e l'altro di 1000 ms. Si è notato che tra una prova e l'altra, per aumentare l'accuratezza, sarebbe meglio aspettare 2/3 minuti e togliere l'alimentazione al motore del braccio robotico per evitarne il surriscaldamento e per allontanarlo di un paio di centimetri dal pilastro, posizionato all'interno della cabina.

10- terminata l'acquisizione di un soggetto, si fa uscire fuori dalla cabina dopo aver recuperato eventuali oggetti personali.

3.2 Protocollo acquisizione video

Il soggetto viene posizionato su una X sul pavimento (che indica la posizione iniziale), l'operatore dà un segnale e inizia l'acquisizione. Viene realizzato un video nel quale il soggetto compie qualche passo per arrivare ad una cassetiera di metallo, presente all'interno del laboratorio, apre un cassetto (il primo), ne estrae un foglio, lo posiziona sulla cassetiera e lo firma. Successivamente, il soggetto torna indietro fino ad arrivare al punto iniziale contrassegnato con la X, il tutto svolto in maniera naturale. Tutto il percorso è registrato mediante cinque telecamere, poste a diverse altezze e diverse orientazioni. Lo scopo è il simulare la sorveglianza stradale o, ad esempio, la videosorveglianza di sicurezza presente negli ATM.

3.3 Protocollo dichiarazione di consenso alla privacy

Le leggi sulla privacy riguardano il complesso di norme che regolano la tutela e l'utilizzo di dati personali. Una violazione riguarda la diffusione di immagini di un soggetto senza il dichiarato consenso di quest'ultimo. Si è avuta la necessità di creare un modulo che permettesse di ottenere la dichiarazione del consenso per l'utilizzo delle immagini dei soggetti sottoposti alla procedura. La soluzione è stata di crearne uno adatto per il caso studio, dato che una ricerca in questo ambito ha portato a moduli molto prolissi e complessi. Per quanto riguarda la compilazione, questa veniva attuata appena il soggetto entrava all'interno del laboratorio dove erano attuate le rilevazioni. Prima di tutto era informato verbalmente in cosa consistesse la procedura e poi era richiesta la compilazione del modulo stesso che poi veniva riposto in un apposito contenitore.

Il modulo in questione è strutturato in quattro parti principali (Figura 3.1):

- 1- compilazione con dati personali del soggetto in questione. Per semplicità sono stati richiesti: il nome e il cognome, il luogo e la data di nascita, la residenza

- 2- “afferma”, dove, in breve, è dichiarato sia la veridicità delle riprese in questione sia l’assunzione di ogni responsabilità per qualsiasi spesa e/o danno derivante dall’utilizzo e dalla diffusione dell’immagine del soggetto e, infine, viene rilasciata la disponibilità di usare le riprese da coloro in possesso di tale modulo.
- 3- “autorizza” nel quale si esprime che la pubblicazione, l’utilizzo e la diffusione può essere attuata con qualunque mezzo senza limitazione spaziale o temporale del materiale audiovisivo e fotografico raccolto in occasione della ricerca scientifica in questione condotta presso l’Università Politecnica delle Marche.
- 4- espressione del consenso alla diffusione del materiale per utilizzarlo senza fini di lucro. In questa sezione viene vietato l’utilizzo in casi che ne pregiudichino il decoro e la dignità personale.
- 5- Infine, viene dichiarata la data in cui viene attuata la rilevazione ed è richiesta la firma del soggetto.

Dichiarazione di consenso alla pubblicazione e diffusione di immagini

1	<p>Il/La sottoscritto/a _____ nato/a a _____ il _____ residente a _____ in via _____ numero civico _____, preso atto dell’informativa ricevuta ai sensi dell’art. 13 del D.Lgs. n. 196/2003, dell’art. 13 Regolamento UE 2016/679, e letta l’informativa dell’Università Politecnica delle Marche sulla protezione dei dati relativa all’uso di immagini, foto e riprese audio-video,</p>
2	<p style="text-align: center;">afferma</p> <ul style="list-style-type: none"> - che le dichiarazioni rilasciate, inclusi dati, informazioni, narrazioni, descrizioni, ricostruzioni, opinioni riguardanti fatti, circostanze, luoghi, eventi e persone da me riferiti, descritti e/o esposti nel corso delle riprese sono vere e non contengono alcun elemento avente natura o finalità pubblicitaria diretta, indiretta o subliminale o che possa provocare lesioni di diritti personali o patrimoniali di terzi; - che sono nella disponibilità di chi scrive i documenti e gli oggetti esibiti, consegnati, in qualunque modo utilizzati, nel corso delle riprese e non ledono i diritti personali ovvero sia patrimoniali di alcuno; - di non aver concesso a terzi diritti confliggenti o in qualche modo altrimenti pregiudizievoli dalla presente dichiarazione; - di assumere ogni responsabilità per qualsiasi pretesa, spesa e danno derivante dall’utilizzo e dalla diffusione della propria immagine o voce, del proprio nome o della violazione di una qualsiasi delle garanzie e dichiarazioni quivi rese; - che nulla osta al rilascio della seguente autorizzazione; <p>e, pertanto:</p>
3	<p style="text-align: center;">autorizza</p> <p>la pubblicazione, l’utilizzo, la riproduzione, la diffusione e la cessione a terzi, con qualsiasi mezzo, senza alcuna limitazione dal punto di vista spaziale e senza limiti di tempo, del materiale audiovisivo e fotografico raccolto in occasione delle/dell’attività di ricerca scientifica condotte/a presso l’Università Politecnica delle Marche ;</p>
4	<p style="text-align: center;">esprime il consenso</p> <p>alla diffusione del materiale audiovisivo e fotografico, per utilizzarle senza fini di lucro.</p> <p>Ne vieta, altresì, l’uso in tutti i casi che ne pregiudichino il decoro e la dignità personale. Ai sensi di quanto previsto dal Regolamento (UE)2016/679 in qualsiasi momento potrà avvalersi del diritto di revocare l’autorizzazione al trattamento del materiale audiovisivo e fotografico.</p> <p>Ancona, _____ in fede</p>

Figura 3.1 - Modulo della privacy consegnato ai soggetti che si sottoponevano alla procedura.

3.4 Acquisizione di un soggetto

In questa parte del capitolo verrà fornito un esempio di acquisizione di un soggetto.

- FOTOSEGNALAMENTO

Dopo aver fatto sedere il soggetto all'interno della cabina e azionato il braccio di posizionamento: come output si avranno 28 foto, quattro per ognuno dei sette angoli selezionati.

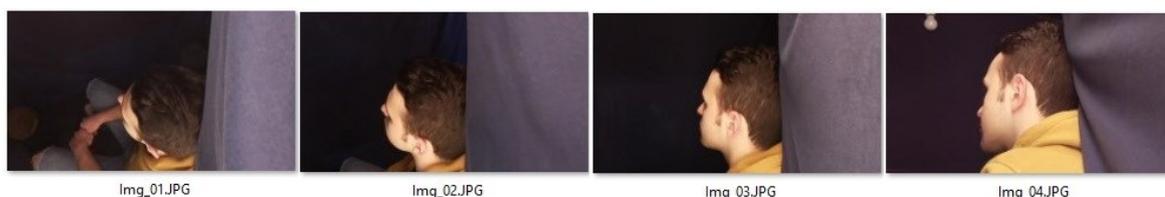


Figura 3.2 - Foto relative al primo angolo.

Come si vede nella Figura 3.2, il braccio di posizionamento inizia la sua corsa da sinistra e nelle prime acquisizioni si può notare la parte posteriore del capo. Si evidenziano le diverse inclinazioni delle telecamere che portano ad avere differenti prospettive dell'immagine. Inoltre, si nota come il soggetto sia posizionato correttamente, infatti l'intero volto è all'interno del riquadro e la testa è appoggiata al rivestimento in gommapiuma posto all'interno della cabina.

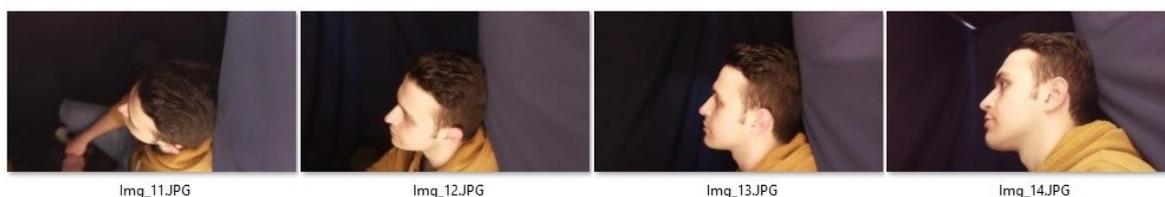


Figura 3.3 - Foto relative al secondo angolo.

Nella Figura 3.3 il braccio robotico è nel secondo angolo di acquisizione, si vede il profilo sinistro del viso. Anche in questo caso si riescono a distinguere le diverse inclinazioni delle *Raspberry Pi Zero W*, in particolare nell'immagine *Img_13* si identifica la telecamera posta ad altezza frontale. Si nota come la telecamera dell'*Img_11* sia posizionata più in alto e dia una visione meno significativa dal punto di vista del fotosegnalamento, ma che può diventare chiave nel riconoscimento di un soggetto a partire da video di videosorveglianza magari stradale.

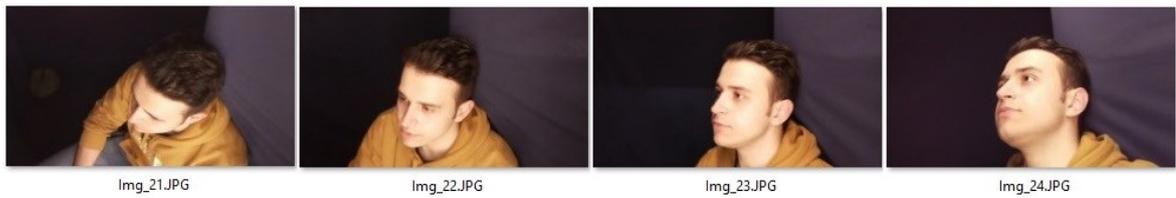


Figura 3.4 - Foto relative al terzo angolo.

Nelle quattro foto illustrate nella Figura 3.4, si vede anche in questo caso il profilo sinistro del viso del soggetto e si inizia ad intravedere il destro. Si evidenzia come la persona guardi un punto fisso davanti a sé e che l'illuminazione e il fuoco delle telecamere sia uniforme.



Figura 3.5 - Foto relative al quarto angolo.



Figura 3.6 – Foto scattata dalla telecamera frontale.

Come si vede nella Figura 3.5, il quarto angolo rappresenta il momento in cui il braccio di posizionamento è posto frontalmente al soggetto. La *Img_33* è l'immagine frontale, sia per la posizione del braccio che per l'orientamento della telecamera. Non in tutti casi l'acquisizione permette di ottenere una foto perfettamente frontale, per questo motivo è stata aggiunta un'altra foto scattata con una telecamera fissa posta frontalmente al soggetto all'interno della cabina (Figura 3.6).

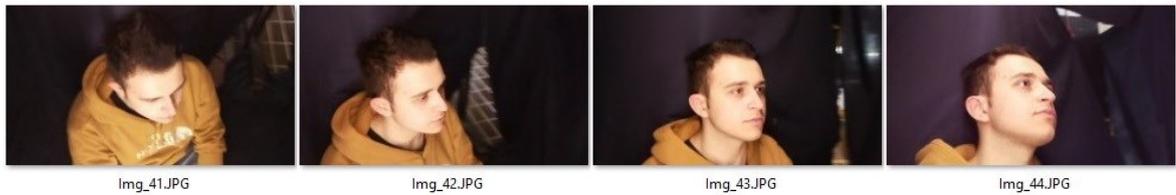


Figura 3.7 - Foto relative al quinto angolo.

La Figura 3.7 rappresenta le acquisizioni attuate dal braccio robotico nella quinta posizione. Le foto mostrano il profilo destro del soggetto in questione, si può notare come si ha una situazione speculare al terzo angolo visionabile in Figura 3.4.

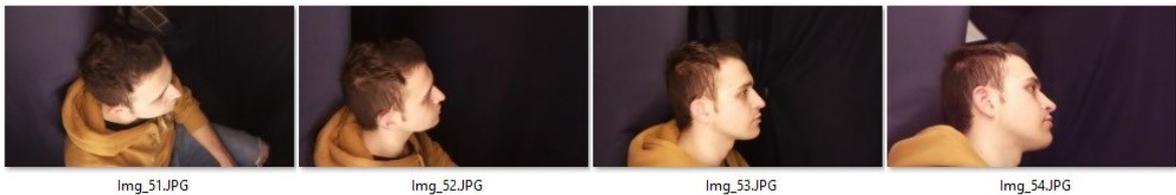


Figura 3.8 - Foto relative al sesto angolo.

Nella Figura 3.8 si ha, allo stesso modo del caso precedente, una specularità con il secondo angolo (Figura 3.3), in questo caso si vede il profilo destro del soggetto.

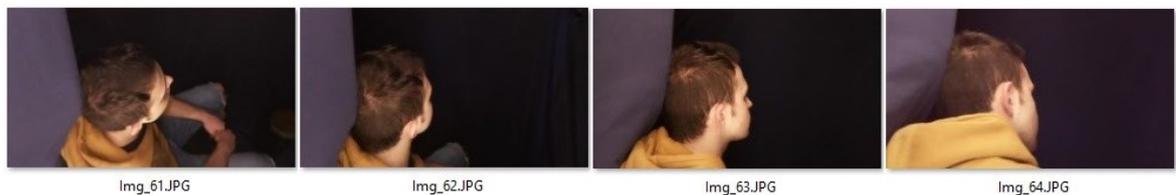


Figura 3.9 - Foto relative al settimo angolo.

Nell'ultima posizione del braccio di posizionamento mostrata nella Figura 3.9, si nota la specularità con il primo angolo (Figura 3.2), infatti è possibile notare la parte posteriore destra del capo.

Guardando la Figura 3.10 che mostra in sequenza tutte le foto acquisite durante il fotosegnalamento, è possibile notare come viene visionato l'output nella sua interezza, ossia come è attuata la scansione del volto.



Figura 3.10 – Output completo del fotosegnalamento.

- VIDEOSORVEGLIANZA

Dopo aver fatto posizionare il soggetto al di sopra della X sul pavimento, che rappresenta il punto di riferimento dal quale iniziare l’acquisizione, si procede con la registrazione della camminata all’interno del laboratorio che mira a una simulazione di un’azioni attuata in ambiente dotato di videosorveglianza. Per riprodurre al meglio la quotidianità sono stati ripresi soggetti con diversi dei loro oggetti personali; nel dataset si riportano persone con occhiali da vista (Figura 3.11), occhiali da sole (Figura 3.16) ma anche berretti.

Nella Figura 3.11 è stato estrapolato un frame del video catturato dalla *Cam 1*. La numerazione delle telecamere è stata scelta a priori ed era indicata su ognuno di esse.

È visionabile la parte sinistra del volto, dato che la telecamera era montata sul supporto in ferro ad un’altezza di circa 1.50 m.



Figura 3.11 – Esempio di un frame del video catturato dalla telecamera 1.

Nella Figura 3.12 è riportato lo screen di un momento del video catturato dalla *Cam 2*. Notando le differenze con la Figura 3.11, si può vedere come l'altezza delle due telecamere sia la medesima, ma il punto di vista è opposto. Infatti, in questo caso, è visibile il profilo destro del viso del soggetto; si ha quindi una prospettiva speculare tra la telecamera 1 e la telecamera 2. Un'altra osservazione è che il frame è più mosso rispetto a quello della Figura 3.11, questo è dato dal fatto che nella prospettiva ripresa dalla *Cam 2* il volto era visionabile in un momento precedente, ossia quando il soggetto in questione era ancora in movimento.



Figura 3.12 – Esempio di un frame del video catturato dalla telecamera 2.

Nel frame estrapolato dalla telecamera 3, visionabile nella Figura 3.13, si nota come la prospettiva sia completamente diversa dalle due precedenti. La *Cam 3* in questione è stata posizionata più in alto (circa 2.10 m) con lo scopo di riprodurre i sistemi di videosorveglianza stradale. Come nella Figura 3.11 è visionabile il profilo destro del soggetto e il frame è più a fuoco rispetto alla Figura 3.12 dato che è riportato il momento in cui il soggetto è intento ad aprire un cassetto della cassetiera in metallo posta all'interno del laboratorio (azione facente parte del protocollo video).



Figura 3.13 – Esempio di un frame del video catturato dalla telecamera 3.

Nella Figura 3.14 è stato estrapolato un frame delle acquisizioni effettuate della telecamera 4.



Figura 3.14 – Esempio di un frame del video catturato dalla telecamera 4.

Quest'ultima è posta ad un'altezza per la quale è facile riprodurre le telecamere di sicurezza presenti all'interno delle ATM. È orientata in modo da catturare il profilo destro del soggetto e, anche in questo caso, è stato scelto un momento dove il soggetto era quasi fermo in modo da riuscire anche a comparare i fotogrammi estrapolati dalla videosorveglianza con la scansione del volto attuata nel fotosegnalamento.

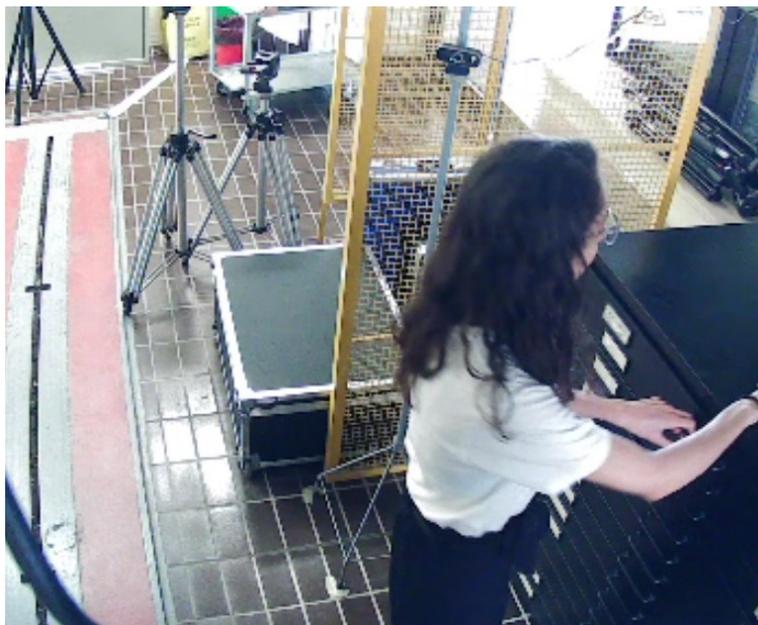


Figura 3.15 – Esempio di un frame del video catturato dalla telecamera 5.

Nell'ultima figura (Figura 3.15) è visionabile un momento del video acquisito dalla telecamera 5. La prospettiva è speculare a quella della telecamera 3 (Figura 3.13), infatti si può notare come in questo caso il volto non è visibile, visto che il soggetto è di spalle. Per quest'ultimo motivo le diverse prospettive sia delle telecamere per la videosorveglianza che nel fotosegnalamento sono un elemento chiave per quanto riguarda il riconoscimento facciale.



Figura 3.16 – Esempio di un frame del video catturato dalle 5 telecamere, il soggetto indossa degli occhiali da sole.

Capitolo 4 Analisi dei risultati

Nella **Errore. L'origine riferimento non è stata trovata.** sono riportati tutti i soggetti facenti parte del dataset. È divisa la parte del fotosegnalamento da quella della videosorveglianza. In quest'ultima i video sono divisi a seconda degli oggetti aggiuntivi che i soggetti hanno durante l'acquisizione. Gli oggetti mirano a coprire parti del volto in modo da mettere alla prova il software di riconoscimento facciale.

In alcune caselle di diversi soggetti è presente un simbolo “-“, che sta ad indicare una mancanza nel dataset delle acquisizioni contrassegnate.

Tabella 4.1 – Elenco di tutti i soggetti del dataset (X = completato).

	FOTOSEGNALAMENTO	VIDEO			
SOGGETTI		NORMALE	CON CAPPELLO	CON BANDANA	CON OCCHIALI DA SOLE
1	X	X			X
2	X	X			
3	X	X			
4	X	X			
5	X	X		X	
6	X	X			X
7	X	X			X
8	X	X			
9	X	X			
10	X	X			
11	X	X			
12	X	X			
13	X	X			X
14	X	X			X
15	X	-			
16	X	X			
17	X	-			
18	X	X			
19	X	X			X
20	X	X			X
21	X	X			X
22	X	X	X		X
23	X	X			X
24	X	X			X
25	X	X			X
26	X	X			
27	X	X			

28	X	X			
29	X	X			
30	X	X			
31	X	X			
32	X	X			
33	X	X			
34	X	X			X
35	X	X			X
36	X	X			
37	X	X			
38	X	X			
39	X	X			X
40	X	X			
41	X	X			
42	X	X			
43	X	X			
44	X	-			
45	X	X			
46	X	X			

Conclusioni

Il dataset proposto in questa tesi presenta 28 foto segnaletiche e cinque video acquisiti da telecamere di sicurezza per 46 soggetti. Il set di dati ha lo scopo di confrontare le tecniche di riconoscimento facciale per l'identificazione dei soggetti nei video utilizzando le foto segnaletiche disponibili [7].

Questi esperimenti hanno portato alle seguenti conclusioni:

- il dataset proposto è adeguato a confrontare le tecniche di riconoscimento facciale. Presenta un'accuratezza minore rispetto a quelli presenti in letteratura che sfruttano solo la foto frontale del soggetto, ma d'altra parte include anche video di videosorveglianza da differenti punti di vista.
- Nella maggior parte dei test si rileva la maggior accuratezza utilizzando la foto frontale e le foto a $+45^\circ$ e -45° .

Sviluppi Futuri

In futuro nel dataset proposto verranno aumentati il numero di soggetti acquisiti, aggiungendo video ad una risoluzione più elevata, in modo da avere maggiore variabilità e costruire un database ancora più rappresentativo della realtà.

Inoltre, potrebbe essere modificata ulteriormente la meccanica del braccio di posizionamento: invece di presentare un movimento continuo, si potrebbe sfruttare il motore passo passo attuando rilevazioni sempre ogni 45° , ma stoppando il braccio robotico in modo da migliorare la messa a fuoco delle foto segnaletiche. Infatti, modificando il metodo di acquisizione, si riuscirebbe ad ovviare il problema dello sfasamento in alcune rilevazioni. Infine, cambiando la meccanica strutturale del braccio di posizionamento, si potrebbe migliorare il problema del surriscaldamento del motore per riuscire ad acquisire più soggetti e più velocemente, senza attendere dei minuti tra una rilevazione e l'altra.

Bibliografia

- [1] KHAIRWA, Anshul, et al. A comprehensive study of various biometric identification techniques. In: 2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12). IEEE, 2012. p. 1-6.
- [2] CONTARDO, Paolo, et al. Deep Learning for Law Enforcement: A Survey About Three Application Domains. In: RTA-CSIT. 2021. p. 36-45.
- [3] MASI, Iacopo, et al. Deep face recognition: A survey. In: 2018 31st SIBGRAPI conference on graphics, patterns and images (SIBGRAPI). IEEE, 2018. p. 471-478.
- [4] PHILLIPS, P. Jonathon, et al. The FERET database and evaluation procedure for face-recognition algorithms. *Image and vision computing*, 1998, 16.5: 295-306.
- [5] KODOSKY, Jeffrey. LabVIEW. *Proceedings of the ACM on Programming Languages*, 2020, 4.HOPL: 1-54.
- [6] ABAS, Kevin; PORTO, Caio; OBRACZKA, Katia. Wireless smart camera networks for the surveillance of public spaces. *Computer*, 2014, 47.5: 37-44.
- [7] Contardo, P.; Sernani, P.; Tomassini, S.; Falcionelli, N; Martarelli, M; Castellini, P.; Dragoni, A. F. The Face Recognition from Mugshots Database: a Dataset to Test the Use of Mugshots from Multiple Points of View for Identification. (manuscript)
- [8] SAMARIA, Ferdinando S.; HARTER, Andy C. Parameterisation of a stochastic model for human face identification. In: *Proceedings of 1994 IEEE workshop on applications of computer vision*. IEEE, 1994. p. 138-142.
- [9] Wolf, L.; Hassner, T.; Maoz, I. Face recognition in unconstrained videos with matched background similarity. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 529–534.
- [10] S. Johnsen and A. Tews, “Real-Time Object Tracking and Classification Using a Static Camera,” *Proc. IEEE Int’l Conf. Robotics and Automation, Workshop on People Detection and Tracking*, 2009.
- [11] Adjabi, I.; Ouahabi, A.; Benzaoui, A.; Taleb-Ahmed, A. Past, Present, and Future of Face Recognition: A Review. *Electronics* 2020, 9, 1188.

Sitografia

Ministero dell'interno:

[https://www.interno.gov.it/it/notizie/sistema-automatico-riconoscimento-immagini futuro-diventa-realta](https://www.interno.gov.it/it/notizie/sistema-automatico-riconoscimento-immagini-futuro-diventa-realta)

Amazon:

https://www.amazon.it/dp/B09BQQV6K3/ref=cm_sw_r_apan_i_dl_29DMA8EKXKEF7PSTNH37?_encoding=UTF8&psc=1

https://www.amazon.it/dp/B08BNM52L8/ref=cm_sw_r_apan_i_CRJ6ABXWVVMG3JRSDFM2?_encoding=UTF8&psc=1

Secom: <https://www.secomitalia.com/>

OBS studio: [Welcome to OBS Studio's documentation! — OBS Studio 27.2.4 documentation \(obsproject.com\)](https://obsproject.com/)

LabView: [LabVIEW | Proceedings of the ACM on Programming Languages](#)

Labteq s.r.l.: [Controllo di accesso ad aree ATM - Labteq](#)

YouTube: [Micro camera per bancomat/ATM IPC0122P3 - YouTube](#)