

POLITECNICO DI MILANO
Facoltà di Ingegneria dell'Informazione
Corso di Laurea Specialistica in Ingegneria Informatica
Dipartimento di Elettronica e Informazione



**MAPPA VISOMOTORIA BASATA SULLA LEGGE DI HERING E
SUL MOTOR BABBLING PER UNA TESTA RIDONDANTE ATTIVA
E UN BRACCIO A 3 GRADI DI LIBERTÀ**

Relatore: Prof. Giuseppina Gini

Correlatore: Ing. Flavio Mutti

Tesi di Laurea di:

Marco ANGIOLETTI

Matr. 749970

Andrea BIANCHI

Matr. 750440

Anno Accademico 2010–2011

Y'know where steel wool comes from? Robot sheep!

Dennis the Menace (Hank Ketcham)

Ringraziamenti

*Ringraziamo il nostro correlatore Mutti Flavio,
per averci seguito e fornito spunti durante la realizzazione di quest'elaborato
e la nostra relatrice e docente Gini Giuseppina,
per la valida istruzione che ci ha offerto e la passione che ci ha trasmesso.*

*Un grazie speciale alle nostre famiglie per il supporto e
in particolare a Manuela, Laura e mamma Lucia per la revisione.*

Marco e Andrea

Sommario

Con questo elaborato mostriamo un controllore bio-ispirato in grado di calcolare il mapping visomotorio di un modello di robot umanoide semplificato. Il sistema fisico è caratterizzato da vincoli meccanici compatibili con caratteristiche umane e composto da un apparato di stereovisione attivo unito ad un braccio robotico, entrambi ridondanti. La testa e le telecamere sono mosse da un controllore biologicamente plausibile basato sulla legge dell'uguale innervazione di Hering. Complessivamente il sistema, grazie alle telecamere, è in grado di percepire una feature nello spazio, calcolare gli angoli dei giunti per fovearla ed eseguire il mapping fra gli angoli di foveazione e la configurazione finale dei giunti del braccio per raggiungere l'obiettivo. Il visuomotor mapping è ottenuto tramite una rete neurale a radial basis functions usando come input gli angoli di foveazione della testa. L'addestramento della rete è eseguito seguendo lo schema del motor babbling utilizzando una Cross-validation a 10-fold per verificarne la robustezza.

Abstract

With this dissertation we propose a bioinspired controller to calculate the visuomotor mapping of a simplified humanoid robot model. The physical system is characterized from human inspired mechanic constraint and composed from an active stereo vision system joined to a robotic arm, both redundant. The active head and the cameras are moved from a controller that draw his inspiration from the Hering's law of equal innervation. The system, thanks to the cameras, is able to perceive a feature in the space, to establish the joint angles to foveate it and then to learn the mapping between the foveation angles and the final configuration of the joints of the arm to reach the objective. The mapping is obtained through a radial basis function neural network using as input the foveation angles. The network training is executed following the motor babbling schema using a 10-fold Cross-validation to confirm his reliability.

Indice

1	Introduzione	1
2	Background	7
2.1	Osservazioni sul termine <i>Biologically inspired</i>	7
2.2	Uomo	9
2.2.1	La retina	9
2.2.2	Il nucleo genicolato laterale	12
2.2.3	La corteccia visiva	12
2.2.4	I campi ricettivi	13
2.2.5	Visione stereoscopica	14
2.2.6	Legge di Hering	18
2.2.7	Area V6A	19
2.3	Robot	24
2.3.1	Visione artificiale	24
2.3.2	I manipolatori	29
2.4	Apprendimento	35
2.4.1	Developmental Learning	35
2.4.2	Reti neurali RBF	41
3	Lavori correlati: visione e reaching	45

4	Sistema	55
4.1	Visione	55
4.1.1	Sistema base	55
4.1.2	Sistema esteso	58
4.2	Reaching	60
4.2.1	Braccio RR	62
4.2.2	Braccio RRR	63
5	Metodi	65
5.1	Visione	65
5.1.1	Sistema base	65
5.1.2	Sistema esteso	67
5.2	Reaching	67
5.2.1	Braccio RR	67
5.2.2	Braccio RRR	74
6	Risultati sperimentali	77
6.1	Visione	77
6.1.1	Sistema base	77
6.1.2	Sistema esteso	82
6.1.3	Confronto fra sistema base e sistema esteso	83
6.1.4	Osservazioni	93
6.2	Reaching	98
6.2.1	Braccio RR	98
6.2.2	Braccio RRR	99
6.3	L'unione di Visione e Reaching	105
7	Conclusioni e sviluppi futuri	107

Elenco delle figure

2.1	Retina	10
2.2	Spazio di Panum	16
2.3	Aree del cervello	21
2.4	Geometria camera pinhole	25
2.5	Coordinate immagine-camera	26
2.6	Rototraslazione sistema di riferimento	26
2.7	Triangolazione	28
2.8	Rette sghembe	29
2.9	Catena cinematica manipolatore	31
2.10	Esempi di workspace	33
2.11	Esempi di ridondanza	34
2.12	Approssimazione con gaussiane	43
2.13	Schema RBF	44
3.1	Distribuzione funzioni Chinellato	52
3.2	Mapping dello spazio di Chinellato	54
3.3	Distribuzione dei centri di Chinellato	54
4.1	Calcolo focal lenght	58
4.2	Sdr senza rotazione collo	59
4.3	Struttura braccio RR	62
4.4	Struttura braccio RRR	63

5.1	Spazio di lavoro braccio RR	68
5.2	K-fold Cross-validation	73
5.3	Spazio di lavoro braccio RRR	74
6.1	Sottospazi di testing	78
6.2	Error map sistema base	80
6.3	Test posizione iniziale sistema base	81
6.4	Error map sistema esteso	84
6.5	Test posizione iniziale sistema esteso	85
6.6	Errore 3D nel piano XZ	88
6.7	Errore 3D nel piano XZ	89
6.8	Distanza fra punto ideale e punto foveato	91
6.9	Confronto sistemi di visione	92
6.10	Confronto sistemi di visione con scatter	92
6.11	Traiettorie telecamere	94
6.12	Campo visivo "alieno"	96
6.13	Test esplorazione 3D "alieno"	96
6.14	Test posizione iniziale "alieno"	97
6.15	Distribuzione centri rete neurale braccio RR	98
6.16	Distribuzione dell'errore di reaching nel braccio RR	99
6.17	Simulazione valori giunti	100
6.18	Scatter del dataset	101
6.19	Direzione dell'errore piani X-Z	102
6.20	Direzione dell'errore piani Z-Y	102
6.21	Distribuzione centri rete neurale braccio RRR	104
6.22	Sistema visomotorio	106

Elenco delle tabelle

4.1	Configurazioni sistema di visione esteso	61
6.1	Confronto risultati	86

Glossario

Body schema	Termine coniato originariamente dal neurologo Sir Henry Head è stato definito come: modello posturale del corpo che attivamente si organizza e modifica
End effector	Parte terminale di un braccio robotico
Feature	Caratteristica
GDL	Grado Di Libertà
Grasping	Movimento di prensione
Panning	Detto anche movimento di pan, è il movimento di rotazione orizzontale di una telecamera
RBF	Radial Basis Function, Funzione a base radiale
Reaching	Movimento con il quale si porta a coincidere l'end effector con un punto nello spazio
SDR	Sistema di riferimento
Target	Obiettivo
Tilting	Detto anche movimento di tilt, è il movimento di rotazione verticale di una telecamera
Visual stream	Canali cerebrali di comunicazione lungo cui viaggiano le informazioni visive

Capitolo 1

Introduzione

La conoscenza e padronanza dei concetti alla base della stereovisione è di fondamentale importanza per lo studio dei sistemi che interagiscono con lo spazio che li circonda, essa permette di sopperire alla mancanza del dato relativo alla profondità di cui soffre la visione monoculare. La percezione precisa della posizione di un oggetto nello spazio 3D abilita il soggetto a nuove e numerose modalità d'interazione con il mondo. Oltre ad una conoscenza più precisa dell'ambiente in cui è calato, potrà perfezionare i propri movimenti sviluppando forme d'apprendimento che prima gli erano precluse. Recenti ricerche dimostrano che la visione negli organismi non si è evoluta per permettere all'organismo di percepire, ma per fornire un sistema di controllo dei movimenti delle terminazioni degli arti [1].

La ricerca negli ambiti della biologia è fonte di numerosi spunti d'ispirazione per la robotica, entrambe queste discipline sono aperte a vicendevoli contaminazioni che ne permettono una crescita virtuosa. L'approccio ingegneristico tende a studiare il sistema naturale per trarre ispirazione e risolvere problemi in maniera analoga alle soluzioni riscontrate in natura [2]. La stereovisione può quindi essere utilizzata come componente fondamentale di un sistema di controllo dell'apprendimento e dei movimenti in un robot. All'aumentare della complessità del sistema di visione control-

lato, si assiste al complicarsi delle equazioni del controllore. La ricerca è stata molto attiva in questo campo negli ultimi 15 anni e sono stati proposti numerosi sistemi di controllo basati su diversi algoritmi di apprendimento e tipologie di reti neurali che permettessero di calcolare la profondità di un punto nello spazio ed apprendere una coordinazione fra direzione dello sguardo e movimento della mano. È in questo contesto che proponiamo un metodo *human inspired*, quindi implementabile su un robot umanoide, per la realizzazione del mapping visomotorio. Il nostro approccio è suddiviso in due fasi, una che deve necessariamente seguire l'altra. Come in natura il raffinamento delle capacità visive ha permesso l'apprendimento di movimenti balistici più precisi degli arti, così nella nostra modellizzazione ad una prima fase di addestramento del sistema visivo, ne seguirà una di apprendimento del mapping visomotorio.

Gli spunti provenienti dalla biologia non si fermano al mantenere per il robot dimensioni e proporzioni delle componenti (braccia e occhi) simili a quelle medie umane, gli schemi biologici ai quali ci ispiriamo sono:

- *Developmental learning*, che a partire da concetti e funzioni innate, o apprese in precedenza, permette di derivare nuovi comportamenti attraverso un approccio incrementale [3], [4], [5], [6] e [7].
- *Motor babbling*, una fase dell'apprendimento tipica dei primi mesi di vita dei bambini volta a sviluppare l'abilità di compiere movimenti precisi degli arti ed esplorare lo spazio peripersonale per il raggiungimento di oggetti [8], [9] e [10].
- *Legge di Hering*, modella il movimento degli occhi sulla base della teoria di uguale innervazione, sostiene l'esistenza di un comando centrale comune per muovere entrambi gli occhi [11].
- *Area V6A*, un'area ben delimitata della corteccia cerebrale che si è dimostrata essere di fondamentale importanza per la trasformazione

dei sistemi di riferimento da coordinate retinocentriche a coordinate dei giunti degli arti [12].

- Rete neurale RBF, uno schema di rete neurale che con la sua struttura tipica ad un solo layer di basis function gaussiane pesate, approssima bene la sinergia che si sviluppa fra i neuroni per generare risposte ben definite a vari tipi di stimoli [13], [14] e [15] .

Con la nostra tesi ci proponiamo quindi i seguenti obiettivi innovativi:

- implementare un sistema di controllo di versione e vergenza bio-ispirato, che quindi permetta la foveazione di un punto rispettando vincoli e comportamenti tipici degli esseri umani. Si verificherà il comportamento delle traiettorie di foveazione gestendo i movimenti oculari con le modalità descritte da Hering.
- Introdurre un giunto di rotazione (panning) nel sistema di visione, precisamente in corrispondenza del collo (sistema esteso) e gestire la ridondanza che esso comporta senza amplificare gli errori e continuando a mantenere traiettorie di foveazione umanoidi.
- Sfruttare la sinergia fra sistema di visione attivo basato sulle leggi di Hering e la rete neurale RBF che esegue il mapping visomotorio, implementando quindi il funzionamento dell'area V6A.
- Completare il sistema introducendo un braccio robotico RRR e gestirne la ridondanza rispettando posture, dimensioni degli arti e angoli dei giunti di matrice umana.
- Risolvere il problema della cinematica inversa tramite l'utilizzo di una mappa visomotoria che associa agli angoli usati per foveare un punto 3D gli angoli del braccio necessari per raggiungerlo, evitando la risoluzione di complicate equazioni cinematiche non lineari.

- Apprendere tramite una rete neurale bio-ispirata, basata sulla radial basis function, una corrispondenza fra: angoli degli occhi e di rotazione del collo, e angoli dei giunti del braccio; assicurandosi che i risultati dell'apprendimento continuino a produrre posture dell'arto paragonabili a quelle dell'uomo.

Per perseguire questi obiettivi abbiamo implementato un modello simulato di robot costituito da un sistema di stereovisione attivo e da un sistema di reaching. L'architettura del sistema è ridondante e caratterizzata da vincoli meccanici di ispirazione umana. È composta, per quanto riguarda la visione, da due telecamere che garantiscono la stereovisione attiva; ad esse è unito un braccio robotico. La testa e le telecamere sono mosse da un controllore bio-ispirato basato sulla legge dell'uguale innervazione di Hering. Globalmente il sistema è in grado di percepire un obiettivo nello spazio, muovere il braccio nello spazio dei giunti per fovearlo e, tramite il motor babbling, eseguire il mappaggio fra gli angoli di foveazione e la configurazione finale dei giunti del braccio per raggiungere la feature. Il visuomotor mapping è ottenuto tramite una rete neurale a radial basis functions che riceve come ingresso gli angoli di foveazione del sistema visivo.

Un'ampia parte della trattazione è dedicata alla validazione dei risultati ottenuti. Al fine di confrontare i due diversi sistemi di visione (base ed esteso) abbiamo realizzato un test di esplorazione dello spazio 3D per studiare l'errore ed un test di foveazione a partire da differenti posizioni iniziali del sistema di visione per verificare la robustezza. L'errore, in base al contesto specifico, è stato misurato sia sullo spazio bidimensionale del *piano immagine*, sia nello spazio extra-personale come distanza euclidea fra il punto effettivamente foveato e quello che si sarebbe dovuto foveare. Per quanto riguarda il sottosistema di reaching, si sono validate le performance del visuomotor mapping appreso dalla rete neurale utilizzando una 10-fold Cross-validation.

Dopo questa breve introduzione, la presente trattazione si svolgerà nel seguente modo:

- nel Capitolo 2 verranno spiegate alcune nozioni fondamentali per la comprensione della tesi riguardanti la visione nell'uomo, concetti base di robotica inerenti ai sistemi di visione e manipolazione, infine si porrà l'attenzione sull'apprendimento nell'uomo e nei sistemi artificiali.
- Nel Capitolo 3 verranno illustrati i lavori, le ricerche e gli esperimenti con i quali ci siamo confrontati, dai quali abbiamo tratto spunto e quelli che costituiscono la base del nostro lavoro, con un focus particolare sul lavoro di Chinellato.
- Nel Capitolo 4 sarà descritta l'architettura del nostro sistema, diviso principalmente nella parte di visione e nella parte di reaching. Ognuno di questi sistemi verrà affrontato in maniera incrementale: partiremo dal sistema di visione base, composto da 3 gdl, per poi estenderlo introducendo una ridondanza nel collo (sistema esteso); partiremo da un braccio robotico RR planare, per poi svilupparlo introducendo una ridondanza nella spalla (braccio RRR).
- Nel Capitolo 5, mantenendo la stessa suddivisione fra visione e reaching, verranno illustrati i metodi che abbiamo utilizzato per implementare i sistemi e risolvere i problemi riscontrati.
- Nel Capitolo 6 si illustreranno i risultati sperimentali del sistema di visione base ed esteso confrontandoli fra loro. Seguiranno i risultati riguardanti le performance del sistema di reaching.
- Nel Capitolo 7 esporremo le conclusioni a cui siamo giunti e gli sviluppi futuri che ci auspichiamo possano derivare da questo elaborato.

Capitolo 2

Background

2.1 Osservazioni sul termine *Biologically inspired*

¹ Per prima cosa è doveroso chiarire cosa si intenda col termine *biologicamente ispirato*. Si possono distinguere tre differenti approcci allo studio e alla ricerca nell'ambito dei sistemi naturali ed artificiali [2]:

- un approccio comparativo (naturale *vs* artificiale), nel quale i sistemi artificiali sono trattati allo stesso modo di quelli naturali e si studiano in modo comparativo i meccanismi e le origini dei fenomeni nei vari sistemi.
- Un approccio modellistico (naturale *via* artificiale), nel quale si usano i sistemi artificiali per studiare il funzionamento di quelli naturali.

¹Le immagini di questo capitolo sono tratte da:

Retina, <http://arching.files.wordpress.com/2012/02/sistema-visivo-umano.pdf>;

Visione Stereoscopica, <http://fc.units.it/ppb/visione/Stereopsi/Stereopsi1.html>;

Struttura del cervello,[12]; Sistemi di visione, [16]; Manipolatori, [17],[18];

Reti neurali, http://www.aiaccess.net/English/Glossaries/GlosMod/e_gm_rbf.htm;

Reti RBF, <http://chrome.ws.dei.polimi.it/images/f/fe/Ke-handout-lecture-nn.pdf>.

- Un approccio ingegneristico (naturale *pro* artificiale), nel quale si studia il sistema naturale per trarre ispirazione e sviluppare soluzioni a problemi ingegneristici analoghi a quelle presenti in natura.

La nostra tesi manterrà un approccio del terzo tipo: ispirandoci all'uomo, al funzionamento del cervello e le strutture neuronali che lo compongono, si svilupperà un sistema di mapping visomotorio per un modello di robot umanoide costituito da un sistema di visione e di reaching. Negli ultimi decenni, riferendosi ad un sistema biologicamente ispirato, si soleva intendere un sistema nel quale, con la presenza di reti neurali, si tentava di ricalcare a basso livello il funzionamento del cervello umano. Parallelamente, grazie ad una più approfondita conoscenza dei sistemi che governano il cervello e dei limiti della robotica, si osserva una tendenza verso una maggiore astrazione rispetto ad una pedissequa applicazione degli schemi biologici. Tipicamente la prassi con la quale si è soliti approcciare l'implementazione di un controllore bio-ispirato per la visione e il reaching, è quella di investigare come le informazioni della rete neurale bio-ispirata possano essere elaborate al fine di controllare i processi di foveazione e raggiungimento di un oggetto.

Il nostro progetto, invece, indaga nella prima parte il modello di controllo sotteso alla rete neurale e compatibile con essa, trascurandone l'implementazione in una rete neurale vera e propria, ed utilizzandolo piuttosto per svolgere ulteriori ricerche riguardo l'introduzione di un grado di libertà aggiuntivo nel sistema grazie al collo (solitamente non implementato nei sistemi di controllo classici).

Nella seconda parte, invece, l'attenzione non viene focalizzata sull'implementazione della rete neurale, che è una RBF standard con MSE, ma sulla gestione della ridondanza del braccio robotico e sul processo di motor babbling, che chiude l'anello del developmental learning. Inoltre il robot è stato simulato con parametri ispirati biologicamente alle misure umane

quali: distanza fra gli occhi, angolo di visione, distanza focale e lunghezza degli arti.

2.2 Uomo

Sebbene un'analisi dettagliata del sistema di visione umana sia trascurabile per la comprensione dei successivi capitoli riguardanti l'implementazione del controllore, riteniamo sia interessante scoprire, almeno ad un livello superficiale e per piacere personale, come viene generato ed analizzato il segnale visivo.

Gli occhi umani sono posizionati ad una distanza media interpupillare di 63 millimetri. La lunghezza focale, cioè la distanza fra il cristallino e il piano sul quale viene messa a fuoco l'immagine del mondo esterno (che in un occhio sano si trova sulla retina) è di 17 millimetri, ed il campo visivo è di circa 170 gradi orizzontalmente e 110 verticalmente. Esaminando l'anatomia del sistema di visione umano, o più in generale, dei mammiferi, si possono distinguere tre unità principali: la retina, il nucleo genicolato laterale e la corteccia visiva. L'insieme di questi tre blocchi costituisce il *visual stream* principale, è attraverso questo percorso che le informazioni provenienti dal mondo esterno sono elaborate sino a giungere alle aree corticali associative dove vengono integrate con le informazioni provenienti da altri canali sensitivi [19].

2.2.1 La retina

La funzione principale dell'occhio è quella di mettere a fuoco sulla retina le immagini percepite dal mondo esterno con la minima distorsione possibile. I fasci luminosi vengono messi a fuoco dalla cornea e dal cristallino e vengono assorbiti dai fotorecettori (bastoncelli e coni) dopo aver attraversato la cavità dell'occhio occupata dall'umor vitreo. La superficie

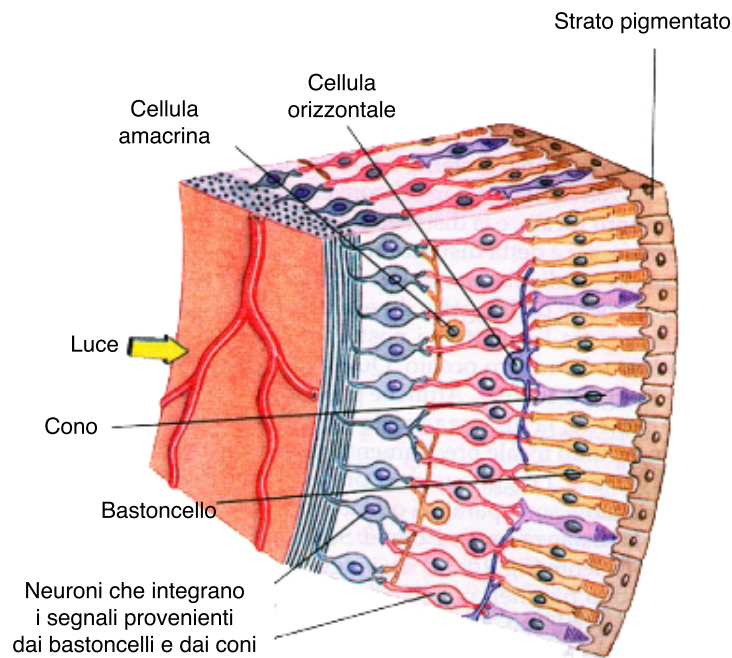


Figura 2.1: Sezione della retina.

posteriore dell'occhio è circondata dall'epitelio pigmentato, al quale aderisce strettamente la retina. Il termine pigmentato deriva dal fatto che le cellule che costituiscono l'epitelio, contengono grandi quantità di melanina conferendogli la tipica colorazione scura; esso ha la funzione di assorbire la luce che non viene fermata dalla retina in modo che non venga riflessa nuovamente su di essa (sezione retina illustrata in Fig. 2.1). I neuroni degli strati prossimali della retina sono amielinici, cioè privi della guaina protettiva di mielina e quindi abbastanza trasparenti, il che permette alla luce di raggiungere i fotorecettori senza venire assorbita o distorta in maniera apprezzabile. Esiste inoltre una zona, chiamata *fovea*, nella quale i corpi cellulari dei neuroni degli strati prossimali della retina sono spostati lateralmente permettendo ai recettori di ricevere l'immagine visiva nella sua forma meno distorta. Questa disposizione è particolarmente accentuata al centro della fovea, nella così detta foveola. È per questo motivo che l'uomo

muove costantemente gli occhi in modo che le immagini si impressionino proprio in questa particolare zona dell'occhio, infatti a soli 10 gradi di distanza dall'asse visivo l'acuità è del 25% in meno.

In condizioni normali gli occhi vengono mossi continuamente e cambiano di direzione circa cinque volte al secondo. Questa attività continua ha un duplice scopo:

- mantenere sulle fovee le immagini degli oggetti in movimento;
- dirigere la fovea sui vari particolari dell'ambiente circostante al fine di ricostruire l'immagine complessiva.

Possiamo distinguere i movimenti oculari in: duzioni, versioni e vergenze. Le *duzioni* sono i movimenti monoculari in qualsiasi direzione. Le *versioni* sono i movimenti binoculari coniugati in cui gli assi visivi mantengono la loro posizione reciproca. Le *vergenze* sono i movimenti binoculari disgiunti, ossia quelli in cui gli occhi si muovono in direzioni opposte. Questa complessa coordinazione di movimenti è gestita da un sistema neuroanatomico molto complesso ed in buona parte ancora ipotetico, il cui scopo principale consiste nel condurre la fovea sul bersaglio e nel mantenerla su di esso per il tempo necessario. Le funzioni elementari che esso svolge sono note ed attribuibili a cinque sistemi neuronali di controllo distinti: il movimento oculare saccadico, il movimento lento d'inseguimento, il riflesso vestibolo-oculare, il movimento optocinetico e il movimento di vergenza. Tutti questi sistemi di controllo utilizzano la stessa via efferente: i motoneuroni dei nuclei oculomotori del tronco encefalico.

La retina umana contiene due tipi di fotorecettori: i bastoncelli ed i coni. I coni sono responsabili della visione a colori ma sono sensibili solo a luci piuttosto intense, i soggetti in cui ne è compromessa la funzionalità sono praticamente ciechi. I bastoncelli assicurano la visione notturna, sono particolarmente sensibili a basse intensità di luce ma non ai colori, funzio-

nano nelle condizioni di luce attenuata presente al crepuscolo e di notte, quando la maggioranza degli stimoli luminosi sono troppo deboli per eccitare il sistema dei cono. Grazie ai fotorecettori le intensità luminose della scena vengono convertite in segnali elettrochimici, fornendo così una prima rappresentazione dell'immagine.

2.2.2 Il nucleo genicolato laterale

Il nucleo genicolato laterale è un porzione del talamo attraverso cui passano quasi tutti gli input sensoriali verso la corteccia ad eccezione di quelli olfattivi, lo chiameremo in seguito *LGN* (dall'inglese *Lateral Geniculate Nucleus*). Gli assoni provenienti dalla retina vanno a connettersi con i neuroni dell'LGN secondo un ordine preciso che dipende dalla posizione dei relativi neuroni sulla retina, in questo modo vengono mantenute le caratteristiche topografiche dello stimolo. Il compito dell'LGN non è ancora del tutto chiaro, a causa delle sue dimensioni e posizione, sembrerebbe svolgere una mera azione di trasferimento, ma potrebbe occupare un ruolo molto più importante. L'esistenza di sinapsi inibitrici e di terminazioni di retroazione che dalla corteccia si innervano nel talamo [20] ci fa supporre che esso abbia una funzione importante nell'integrazione spazio-temporale dell'informazione. Dai neuroni dell'LGN partono le fibre nervose dirette alla corteccia cerebrale, dove viene perfezionata l'elaborazione dei segnali visivi rendendo possibile la percezione cosciente.

2.2.3 La corteccia visiva

La corteccia cerebrale è divisa in un certo numero di aree, quelle che si occupano dell'elaborazione del segnale visivo possono essere raggruppate in un sottoinsieme definito *corteccia visiva*. La corteccia visiva può essere organizzata gerarchicamente e ciascuna area presenta una mappa retinoptica più o meno precisa. La prima area della corteccia visiva è l'*area striata*: qui

si innervano le terminazioni provenienti dal LGN e viene rappresentato in modo abbastanza dettagliato circa metà del campo visivo, ciascuna porzione dell'area striata risponde a proprietà locali di una piccola zona di esso. Salendo nella gerarchia, il mapping diventa meno definito e nello stesso tempo i neuroni rispondono a caratteristiche sempre più complesse dello stimolo visivo: ci sarà un'area sensibile prevalentemente al movimento, un'altra al colore e così via. Progressivamente il confine fra le aree percettive e quelle associative diventa sempre più labile, un neurone delle aree in posizione più elevata nella gerarchia non è in grado di stabilire la posizione nel campo visivo da cui proviene lo stimolo e risulta sensibile a caratteristiche così complesse di esso da rendere impossibile una caratterizzazione individuale della cellula [20]. In seguito, parlando di aree corticali, ci riferiremo esclusivamente a quelle sensoriali ed in particolare all'area striata. Possiamo dividere la corteccia visiva in almeno sei strati, basandoci sulla disposizione delle cellule su piani paralleli alla superficie corticale. Gli strati corticali sono accoppiati fra loro tramite una fitta rete di connessioni verticali, le afferenze dal LGN terminano soprattutto nel quarto strato, dal secondo al sesto strato viene prodotta invece la maggior parte dei segnali in uscita. Queste, insieme ad altre connessioni retroattive, giocano un ruolo fondamentale nell'elaborazione corticale dell'informazione.

2.2.4 I campi ricettivi

Il grande numero di collegamenti fra i neuroni rende impossibile dedurre in che modo le serie successive di neuroni procedono all'elaborazione delle informazioni visive; si dovrebbe conoscere il significato di ogni singolo segnale elettrico prodotto da ogni singola cellula per effettuare uno studio dei tessuti neurali. Porsi il problema di quale sia il significato del segnale elettrico prodotto da ogni cellula, non è impreciso quanto sembra, in realtà corrisponde all'osservare quale specifico evento del mondo esterno

provochi un'attivazione di quel neurone. Il problema viene così riformulato introducendo il concetto di campo ricettivo. Il campo ricettivo indica la regione dell'area visiva in cui deve avvenire un determinato evento per stimolare o inibire il neurone interessato. Per determinare i profili dei campi ricettivi si procede sperimentalmente misurando l'attività elettrica dei neuroni mediante un microelettrodo mentre vengono presentati vari tipi di stimoli visivi al soggetto in diverse aree del campo visivo. La risposta dei neuroni non è uniforme e ad eccezione dei fotorecettori e degli elementi delle prime stazioni retiniche, la luce diffusa a tutto il campo visivo rappresenta uno stimolo inefficace. Generalmente, quindi, il campo ricettivo non è omogeneo, ma è caratterizzato da sottoregioni ON e OFF la cui stimolazione luminosa separata ha effetti opposti sulla risposta cellulare. La stimolazione della sottoregione ON con un fascio luminoso provoca un aumento nell'attività elettrica della cellula, mentre uno stimolo scuro ne provoca l'iperpolarizzazione. Del tutto complementare è il comportamento del neurone a seguito della stimolazione di una sottoregione OFF del suo campo ricettivo: il buio accresce la sua attività spontanea, mentre la comparsa della luce inibisce la risposta. Tenendo conto di queste considerazioni, un campo ricettivo può essere modellato da una funzione bidimensionale che assumerà valori positivi in corrispondenza delle regioni ON e negativi in corrispondenza di quelle OFF. I profili così ottenuti caratterizzano nello spazio le proprietà locali dei neuroni a cui si riferiscono.

2.2.5 Visione stereoscopica

La visione stereoscopica permette all'uomo di percepire la profondità. Nonostante un punto venga fissato da due occhi e quindi si formino due immagini separate e leggermente diverse di quell'oggetto sulle retine, esso viene visto come uno solo. Sebbene un punto posto nella metà destra del campo visivo binoculare, formi la sua immagine sulla parte nasale della

retina dell'occhio destro e su quella temporale (quella più vicina alla tempia) dell'occhio sinistro, esso viene percepito come singolo e localizzato a destra del campo visivo [19]. Ogni elemento retinico stimolato dall'immagine di un oggetto ha un definito valore spaziale (segno locale) e quindi, un'impressione visiva, oltre a determinare luce e colore, ha una sua localizzazione in una determinata direzione nello spazio visivo (legge dell'identica direzione visiva di Hering). L'immagine di un oggetto che si forma nei due occhi viene percepita come unica in quanto in ogni occhio a livello retinico esiste un punto che ha lo stesso valore spaziale di un altro punto situato nella retina dell'occhio controlaterale. Questi elementi retinici, non simmetrici dal punto di vista anatomico, ma accoppiati per comune direzione visiva (e quindi per percezione spaziale) vengono definiti punti retinici corrispondenti e stanno alla base del meccanismo della fusione.

Possiamo definire nel campo visivo un'area detta *area di Panum*. Quest'area ha una forma curva più stretta nella zona del punto di fissazione dello sguardo e progressivamente più ampia tanto quanto più ci si porta alla periferia; deriva da questo che una maggiore disparità è comunque compatibile con una visione unica alla periferia più che nelle parti centrali del campo visivo (Fig. 2.2). Di un oggetto incluso nell'area di Panum, l'occhio destro vede maggiormente la parte destra, mentre quello sinistro la parte sinistra. Pertanto, le due immagini retiniche sono lievemente diverse e cadono su aree che non sono perfettamente corrispondenti ma che vengono ugualmente fuse in una sola conferendo all'oggetto impressione di solidità e senso di profondità. La stereopsi risulta quindi dalla lieve disparità con cui sono visti gli oggetti purché siano compresi nell'area di Panum. Al di fuori di tale area, un punto-oggetto viene percepito doppio perché la sua immagine si forma nelle due retine su punti retinici non corrispondenti (punti retinici disparati).

La distanza e l'angolazione con cui l'oggetto viene fissato (per distanze

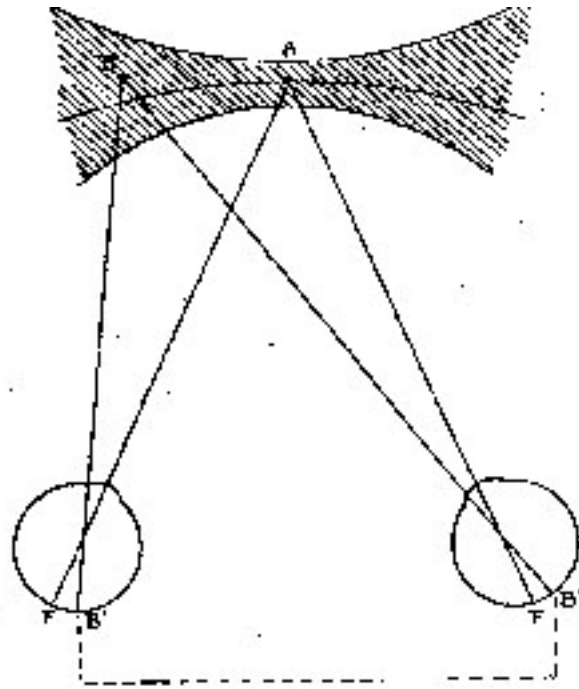


Figura 2.2: Spazio di Panum.

inferiori ai 30 metri) non sono perfettamente uguali nei due occhi: in condizioni normali l'immagine originata dalla fissazione di un oggetto cade a livello foveale grazie ai movimenti di vergenza: poiché gli occhi distano l'uno dall'altro circa 6 centimetri, qualsiasi oggetto che si trovi più vicino o più lontano rispetto al punto di fissazione proietta l'immagine ad una certa distanza dalla fovea; gli oggetti più vicini proiettano la loro immagine su punti della retina più distanti in senso orizzontale e viceversa. La distanza fra immagine del punto fissato ed immagine dell'altro punto prende nome di *disparità retinica*; il sistema visivo è capace di calcolare tale disparità e di assegnare un senso di maggiore o minore profondità agli oggetti dello spazio visivo. Pertanto la percezione delle immagini retiniche è dotata di una leggera diversità che sta alla base della disparità relativa e che nella fusione costituisce la base della percezione della profondità. Il valore stereoscopico degli elementi retinici strettamente corrispondenti è pari allo zero,

mentre la massima sensibilità della stereopsi si ha nelle loro immediate vicinanze. È necessario effettuare una distinzione fra la disparità fine, che da luogo alla percezione tridimensionale stabile (stereopsi manifesta), e quella grossolana che serve solo ad indirizzare nel verso giusto i movimenti di convergenza degli occhi (stereopsi qualitativa). Il senso stereoscopico non dipende soltanto dalla visione binoculare (in cui i fattori principali, almeno per gli oggetti posti a distanza ravvicinata, sono rappresentati dalla convergenza e dall'accomodazione); ne è la prova il fatto che anche soggetti monoculari possono avere il senso della profondità (stereopsi secondaria, che agisce da sola anche nei soggetti binoculari quando si fissano oggetti ad una distanza superiore ai 30 metri, dato che a tale distanza si assume che i raggi luminosi siano pressoché paralleli). In tal caso, molteplici sono gli elementi monoculari empirici che giocano un ruolo nel concorrere alla formulazione del giudizio della distanza relativa degli oggetti:

- Il movimento parallattico: la velocità di spostamento di un oggetto vicino sembra maggiore di quella di uno lontano.
- La prospettiva lineare: un oggetto di grandezza costante sottende angoli progressivamente minori man mano che ci si allontana. L'esempio più classico è quello di due linee rette parallele che sembra tendano a convergere con la distanza: tanto maggiore è la convergenza, tanto maggiore è la distanza degli oggetti nella regione della convergenza stessa.
- La sovrapposizione dei contorni: un oggetto che interrompa i contorni di un altro viene percepito dall'apparato visivo come anteposto all'altro.
- La distribuzione delle luci e delle ombre: il chiaro-scuro genera una impressione del rilievo e quindi di profondità relativa.

- La familiarità con gli oggetti noti: la distanza fra due oggetti di dimensioni note viene giudicata in base alla loro grandezza apparente.
- La prospettiva aerea: è stato dimostrato che l'atmosfera influenza il contrasto ed il colore degli oggetti situati più lontano.

2.2.6 Legge di Hering

I movimenti oculari per fissare e seguire volontariamente un oggetto, si sono evoluti negli animali di pari passo con l'evoluzione della fovea. I movimenti saccadici (rapidi movimenti volontari degli occhi), permettono allo sguardo di muoversi rapidamente da un lato all'altro della scena visiva. Un movimento combinato di entrambi gli occhi nella stessa direzione è chiamato versione. La versione può essere orizzontale, verticale o in direzione obliqua. Un movimento combinato di entrambi gli occhi in direzioni opposte, ma con angolo uguale, è chiamato vergenza. L'angolo della vergenza, insieme con la versione e l'angolo di tilting, descrivono in maniera univoca la posizione del punto che si sta fissando.

Negli umani e negli altri primati, quando uno stimolo visivo cambia direzione e distanza simultaneamente, il movimento degli occhi è il risultato del controllo combinato di versione e vergenza. Questo fenomeno è spiegato dalla legge di Hering, il quale sostiene che gli occhi siano egualmente innervati e che questa connessione dei muscoli oculari sia il motivo per il quale gli occhi si muovono in maniera sincrona. Anche i movimenti apparentemente monoculari possono essere visti come la composizione dei movimenti coniugati di versione e vergenza.

Più precisamente il movimento può essere diviso in tre fasi che coinvolgono due componenti indipendenti: per prima una breve fase durante la quale inizia la vergenza, un rapido movimento saccadico e un graduale completamento della variazione simmetrica della vergenza necessa-

ria a fissare l'obiettivo. La legge di Hering fornisce quindi un meccanismo strutturale per la coordinazione di entrambi gli occhi.

2.2.7 Area V6A

Negli esseri viventi dal sistema nervoso più semplice, risulta evidente che le capacità percettive dell'organismo si sono evolute in stretta relazione col raggiungimento di specifici obiettivi motori.

Vision did not evolve to enable organisms to perceive. It evolved to provide distal control of their movements...

[Goodale 2000, page 365]

Gli esseri umani e altri primati possiedono una capacità superiore nel rapportarsi con gli oggetti che li circondano; recenti ricerche hanno dimostrato che le aree della corteccia parietale posteriore dei primati sono in grado di rappresentare l'ambiente circostante in una maniera utile al compimento di movimenti coordinati complessi di braccia e occhi. Sembra che i sistemi di riferimento necessari a codificare le informazioni visuali e quelli necessari per pianificare e monitorare i movimenti degli arti siano mantenuti contestualmente dalla stessa popolazione di neuroni.

La corteccia visiva del cervello dei primati è organizzata in due canali paralleli chiamati *dorsal stream* e *ventral stream*. Il primo elabora i dati visivi con l'obiettivo principale di dotare il soggetto dell'abilità di interagire con l'ambiente, solitamente sintetizzato col termine "visione per l'azione". Il secondo è dedicato al riconoscimento degli oggetti e ai processi concettuali "visione per la percezione". L'area dorsale è quella più strettamente correlata alle azioni di raggiungimento e grasping, nell'area V6A e circostanti vengono eseguite le trasformazioni dal sistema di riferimento retinico a quello della mano grazie alla codifica simultanea dello spazio peripersonale in diversi sistemi di riferimento.

Per dare un'immagine più ampia si illustrano ora le aree più direttamente coinvolte durante l'interazione di un soggetto con il suo spazio peripersonale: i dati visivi, nei primati, fluiscono dalla retina ai nuclei genicolati (LGN) del talamo e poi prevalentemente alla corteccia visiva primaria (V1) nel lobo occipitale. I due cammini visivi principali vanno da V1 e la vicina V2 alla corteccia parietale posteriore (PPC) ed inferiore (IT). Le informazioni riguardanti gli oggetti fluiscono attraverso il percorso ventrale attraverso V3 e V4 al complesso occipitale laterale (LOC), che si occupa del riconoscimento degli oggetti. L'area dorsale può essere ulteriormente suddivisa in due percorsi paralleli che si occupano uno dei movimenti in prossimità degli oggetti (i.e. reaching) e l'altro del movimento delle articolazioni quando gli oggetti sono più distanti (i.e. grasping). Il percorso dorso-mediano dedicato al raggiungimento degli oggetti include l'area visiva V6, l'area visomotrice V6A e l'area mediana interparietale (MIP). Per quanto riguarda il grasping, le informazioni legate agli oggetti fluiscono attraverso il percorso dorso-laterale, includendo l'area V3A e l'area caudale intraparietale (CIP), poi raggiungono il solco anteriore intraparietale (AIP), l'area dedicata al grasping nel cervello dei primati che le proietta principalmente all'area ventral promotor (PMv). I piani motori calcolati da PMd e PMv sono inviati alla corteccia motrice primaria (M1) che rilascia un segnale d'azione appropriato [12], si veda Fig. 2.3.

Le aree più strettamente legate al movimento di raggiungimento di un oggetto sono l'area corticale V6A e MIP, entrambe ricevono come ingresso principale i segnali dell'area V6 e li inoltrano alla corteccia dorsale premotoria. Per elaborare un'azione adeguata su un obiettivo esterno, il fascio dorsale, attraverso i suoi due sottofasci paralleli, gestisce contemporaneamente le informazioni riguardanti l'oggetto provenienti dalla retina e quelle riguardanti i dati i propriocettivi riferiti ad occhi, testa e mani. L'area V6A rappresenta un nodo fondamentale di questa complessa rete. Infatti

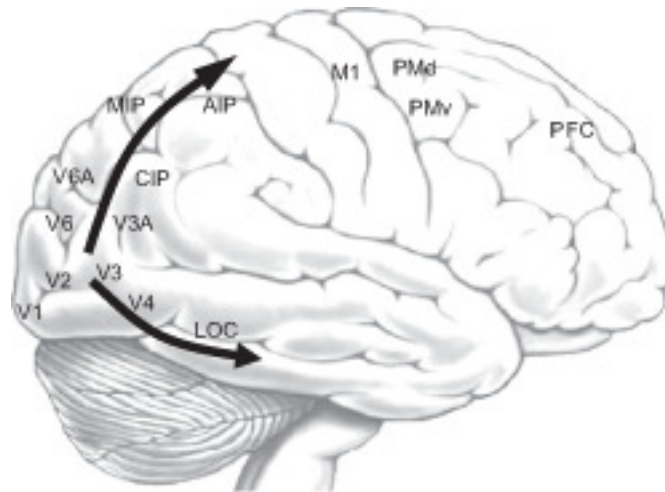


Figura 2.3: I due percorsi delle informazioni visive nel cervello umano: sopra stream dorsale, sotto stream ventrale. Entrambi coinvolti nel reaching e nel grasping.

utilizza le informazioni riguardanti la posizione degli occhi e la direzione dello sguardo per stimare la posizione degli oggetti circostanti e guidare i movimenti per raggiungerli.

Nell'area V6A sono stati trovati due tipi di neuroni che permettono di sostenere quest'ipotesi. Il campo ricettivo dei neuroni del primo tipo è organizzato in coordinate retiniche, ma riescono a codificare posizioni spaziali grazie alla modulazione dello sguardo. Il campo ricettivo dei neuroni del secondo tipo è organizzato secondo la disposizione reale assoluta dello spazio peripersonale del soggetto. In aggiunta, l'area V6A contiene dei neuroni che rappresentano l'obiettivo da raggiungere in coordinate retino-centriche ed altri che utilizzano coordinate spaziali. Queste osservazioni suggeriscono un ruolo critico dell'area V6A nella trasformazione graduale da coordinate retino-centriche ad un sistema di riferimento centrato nella mano. Inoltre, alcuni neuroni dell'area V6A sono direttamente coinvolti nell'esecuzione dei movimenti di raggiungimento e prensione di un oggetto, implicando che quest'area si occupa di effettuare le trasformazioni visomotorie richieste per il controllo volontario del movimento dei giunti

del braccio, integrando segnali visivi, propriocettivi e motori per raggiungere e orientare la mano secondo un obiettivo dato nello spazio 3D.

Esperimento: dall'attivazione neurale all'interazione sensomotora

Le ricerche riportate da Chinellato in [12] e [21] rappresentano una prova tangibile della presenza dei due tipi di neuroni nell'area V6A del cervello, della quale abbiamo parlato poco fa. Chinellato analizza l'esperimento realizzato all'Università di Bologna su due macachi. Durante l'esperimento vengono raccolti dei dati mentre le scimmie compiono due diversi tipi di azioni col braccio, accomunate dal mantenere lo sguardo fisso verso un obiettivo. Le scimmie compiono i movimenti al buio (in modo da ridurre il numero di stimoli), con la testa immobilizzata e gli occhi orientati verso un punto preciso nello spazio detto *fixation point*. L'esperimento è organizzato in quattro fasi: *fix*, *delay*, *move* e *hold*, durante le quali la scimmia deve muovere la mano da un pulsante Inizio ad una posizione Target indicata da un led (fra i tre disponibili) posto esattamente di fronte a lei, ed ha inizio quando la scimmia decide di premere un pulsante vicino al suo petto.

Il *constant reaching protocol* si svolge nel seguente modo:

- **Fix:** la scimmia preme il bottone Inizio. Dopo 1000ms al buio, uno dei tre led (definito *fixation led*) si accende in verde. Esso può coincidere con l'obiettivo o essere alla sua sinistra o destra di 3.7cm.
- **Delay:** la scimmia era tenuta a fissare il led acceso senza muoversi attendendo che il colore del led cambiasse 1000-1500ms.
- **Move:** allo scadere del tempo la scimmia deve premere il bottone Target.
- **Hold:** il pulsante Target deve essere mantenuto premuto finché il *fixation led* non si spegne, per poi riportare il braccio alla posizione iniziale.

Durante il *foveal reaching protocol* le quattro fasi sono analoghe al constant reaching protocol, ma il fixation led si accende sempre in corrispondenza del pulsante Target e quindi la scimmia segue con lo sguardo il movimento della sua mano.

Una volta raccolti i dati dell'esperimento i neuroni vengono classificati secondo la loro selettività, ad esempio se rispondono preferibilmente ad una particolare condizione (posizione del fixation led) di una singola fase. Si è osservato che il cambio di protocollo si ripercuote principalmente sulla componente motoria della risposta neurale, mentre l'orientamento dello sguardo (fix) non è influenzato dal cambio di movimento. Nelle fasi delay e hold si ha invece in entrambe una risposta formata sia da componenti motorie che visuali. Inoltre, nonostante la fase delay possa sembrare dominata solamente da risposte di tipo visuale, in realtà si ha anche l'attivazione di neuroni motori; è quindi lecito supporre venga attivato un processo di pianificazione del movimento. I neuroni attivati nelle fasi di delay e hold sia del constant protocol che del foveal sembra siano in grado di codificare sia la direzione dello sguardo che il raggiungimento dell'obiettivo, anche se diversi. È interessante inoltre osservare come le componenti principali nelle fasi delay e move rimangano consistenti fra i due protocolli; non variando la posizione dell'obiettivo nei protocolli, questo risultato sembra confermare la predominanza in queste fasi dei neuroni motori.

L'analisi suggerisce quindi sia che i neuroni dell'area V6A rimangono attivi anche mentre vengono compiuti i movimenti per fornire un controllo ad anello nella rete feed-forward, sia l'esistenza di una popolazione mista di funzioni a base radiale con una selettività differente per le diverse fasi dell'azione nello spazio visomotorio. Questi neuroni sono quindi dei buoni candidati per compiere le trasformazioni dei sistemi di riferimento e mantenere una rappresentazione sensomotoria integrata dello spazio visuale.

2.3 Robot

2.3.1 Visione artificiale

La visione è un'abilità sensoriale con una grande potenzialità, la capacità di percepire informazioni sull'ambiente circostante è così importante che fin dagli albori della robotica si è pensato di dotare di funzionalità visive i sistemi di intelligenza artificiale [16].

Il modello della telecamera

La formazione dell'immagine è il processo in cui un sensore visivo cattura gli stimoli luminosi e i loro cambiamenti per formare un'immagine bidimensionale. Per modello geometrico della formazione dell'immagine si intende la relazione che intercorre fra la posizione di un punto nella scena e la corrispondente posizione del punto nel piano immagine. Il modello più semplice e comune della fotocamera è il cosiddetto modello *pinhole*. Quando si usa un modello *pinhole*, la mappatura geometrica fra 3D e 2D è detta *proiezione prospettica*. Il centro della proiezione, cioè il punto in cui tutti i raggi si intersecano, è detto *centro ottico* o centro della telecamera C e dista f , *lunghezza focale*, dal piano immagine; la linea passante per C e perpendicolare al piano immagine è detta *asse ottico* e il punto di intersezione fra asse e piano immagine è detto *punto principale* p . In questo modello, ogni punto dello spazio con coordinate $X = (X, Y, Z)^T$ è mappato nel punto del piano immagine dove la linea congiungente il punto X con il centro C interseca il piano (Fig. 2.4). È facile quindi calcolare che il punto X è mappato nel punto $(f \frac{X}{Z}, f \frac{Y}{Z}, f)$ del piano immagine. Si ha quindi, ignorando l'ultima coordinata, il mapping da \mathbb{R}^3 a \mathbb{R}^2

$$(X, Y, Z)^T \rightarrow (f \frac{X}{Z}, f \frac{Y}{Z}). \quad (2.1)$$

Se rappresentiamo i punti della scena e quelli del piano immagine con vettori omogenei, la proiezione può facilmente essere espressa come una tra-

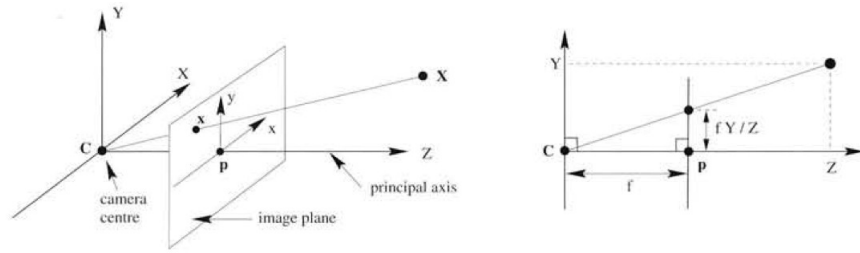


Figura 2.4: Geometria del modello pinhole. C è il centro della telecamera e p il punto principale. Il centro della telecamera in questo caso è posto nell'origine del sistema di riferimento.

sformazione lineare di coordinate omogenee e può essere scritta in forma matriciale:

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} fX \\ fY \\ Z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (2.2)$$

che può essere abbreviata in

$$x = PX \quad (2.3)$$

con X coordinate omogenee del punto 3D, x rappresentazione del punto nell'immagine e P matrice 3×4 detta *matrice di proiezione* della telecamera.

Se il punto principale $p = (p_x, p_y)$ non coincide con l'origine delle coordinate del piano immagine (Fig. 2.5), è necessario introdurre nel mapping una traslazione

$$(X, Y, Z)^T \rightarrow \left(f \frac{X}{Z} + p_x, f \frac{Y}{Z} + p_y \right) \quad (2.4)$$

che in coordinate omogenee diviene

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} fX + Zp_x \\ fY + Zp_y \\ Z \end{bmatrix} = \begin{bmatrix} f & p_x & 0 \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \quad (2.5)$$

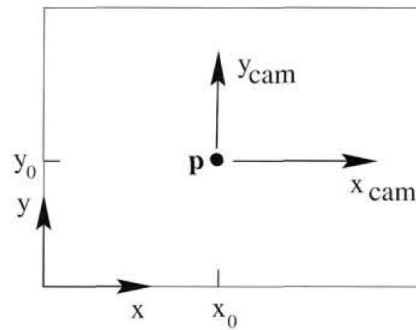


Figura 2.5: Sistema di riferimento dell'immagine (x, y) e della telecamera (x_{cam}, y_{cam}) .

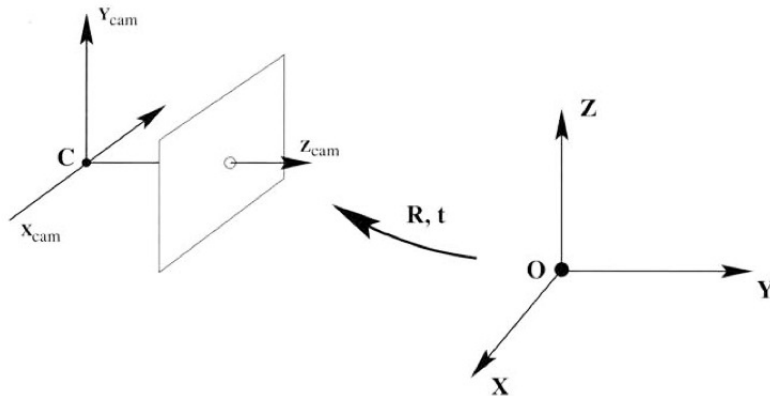


Figura 2.6: Trasformazione euclidea fra il sistema di riferimento del mondo e quello della telecamera.

Ponendo

$$K = \begin{bmatrix} f & p_x \\ & f & p_y \\ & & 1 \end{bmatrix} \quad (2.6)$$

si può scrivere

$$x = K [I | 0] X. \quad (2.7)$$

K è detta *matrice di calibrazione* della telecamera e contiene i suoi parametri intrinseci come, per esempio, la distanza focale.

Solitamente le coordinate dei punti dello spazio non sono rappresentate nel sistema di riferimento della telecamera ma nel sistema di riferimento

del mondo. È possibile passare da un sistema all'altro attraverso operazioni di rototraslazione (Fig. 2.6). La posizione e l'orientamento della telecamera nel sistema di riferimento mondo, detti parametri estrinseci, sono definiti da un vettore C , coordinate del centro ottico, e da una matrice 3×3 di rotazione R . Dette X le coordinate di un punto nel sistema di riferimento mondo e X_{cam} lo stesso punto nel sistema di riferimento della telecamera, si può scrivere che

$$X_{\text{cam}} = \begin{bmatrix} R & -RC \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} R & -RC \\ 0 & 1 \end{bmatrix} X. \quad (2.8)$$

Unendo questa equazione con la (2.7) si ottiene

$$x = KR [I \mid -C] X. \quad (2.9)$$

Ponendo $t = -RC$ la matrice della telecamera diviene

$$x = K [R \mid t]. \quad (2.10)$$

La determinazione dei parametri intrinseci ed estrinseci viene effettuata attraverso l'operazione di *calibrazione*. Poiché questi parametri regolano il modo in cui un punto 3D viene proiettato nel piano immagine, l'idea è che conoscendo le coordinate di punti della scena noti a priori (punti di calibrazione) sia possibile risalire ai valori incogniti dei parametri risolvendo le equazioni di proiezione prospettica.

La stereovisione

La trasformazione prospettica che mappa un punto della scena nel piano bidimensionale di una telecamera implica la perdita dell'informazione relativa alla profondità. Due punti distinti dello spazio, intersecati dalla stessa retta che parte dal centro ottico vengono infatti mappati sullo stesso punto nel piano immagine. Un modo per ovviare a questo problema è

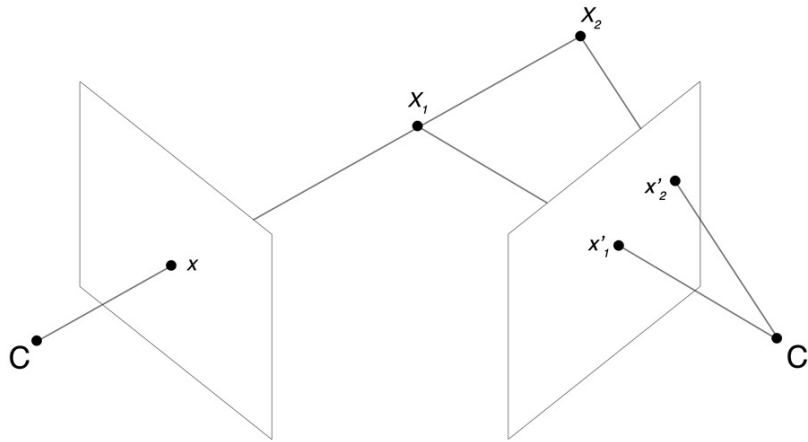


Figura 2.7: Le proiezioni nel piano immagine dei punti X_1 e X_2 sono coincidenti quando viste dalla telecamera di centro C . Tuttavia, tali punti, visti da un'altra telecamera di centro C' , non collineare a C , hanno due distinte proiezioni x'_2 e x'_1 .

l'uso di due o più telecamere, come succede nell'uomo: una delle facoltà del nostro cervello è quella di poter fondere le due immagini retiniche provenienti dagli occhi in un'unica immagine, dando senso di profondità e tridimensionalità. Il principio alla base della visione stereoscopica consiste in una *triangolazione* mirata a mettere in relazione la proiezione di un punto della scena con i piani immagine delle telecamere che compongono il sistema di visione stereoscopica. La triangolazione consiste nell'intersezione fra i due raggi ottici propagati dai centri ottici e passanti nelle proiezioni nei piani immagine (Fig. 2.7). È possibile effettuare analiticamente questo procedimento, ossia risalire alla posizione di un punto X nello spazio a partire dalle coordinate retiniche nelle due telecamere, $x = PX$ e $x' = P'X$. Queste equazioni possono essere combinate in un sistema lineare in X del tipo $AX = 0$, facilmente risolvibile con metodi di calcolo numerico.

Il metodo di triangolazione lineare è applicabile solo nel caso ci si trovi in simulazione, nella realtà gli errori presenti nei punti calcolati x e x' faranno in modo che i raggi siano sghembi. Questo significa che non c'è un punto

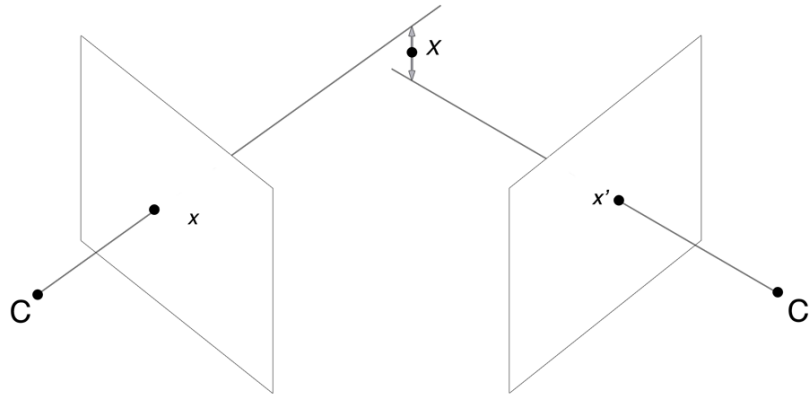


Figura 2.8: I raggi proiettati dalle misure imperfette dei punti x e x' solitamente sono sghembi nello spazio 3D.

X che esattamente soddisfa $x = PX$ e $x' = P'X$ (Fig. 2.8). In tal caso metodi basati sulla minimizzazione della distanza fra i due raggi permettono di stimare con data precisione la posizione di X .

2.3.2 I manipolatori

Un robot manipolatore è una catena cinematica sequenziale aperta composta da corpi rigidi (*link*) uniti da *giunti*. Un'estremità della catena è definita base e solitamente è fissata a terra, mentre l'ultimo link è terminato con uno strumento di lavoro, detto *end effector*.

Esistono due tipi di giunti a seconda del tipo di moto. Il giunto *rotoidale* R permette un moto relativo rotatorio intorno all'asse di rotazione, mentre il giunto *prismatico* T realizza un moto relativo di traslazione fra i due bracci. Non c'è un'analogia antropomorfica per il giunto prismatico. Le variabili di giunto sono i parametri che possono cambiare in un giunto: un angolo, θ , per il giunto rotoidale e la lunghezza, d , per il giunto prismatico. Ogni giunto è un grado di libertà (*gdl*) del robot. Per raggiungere i punti in uno spazio tridimensionale il manipolatore deve avere almeno tre *gdl*,

distribuiti nel braccio, mentre per avvicinarsi con ogni orientamento ne servono altri tre, solitamente distribuiti nel polso. Se il numero di gradi di libertà è maggiore del numero di gradi di mobilità, il robot si dice *ridondante*.

La cinematica

Per dati valori di variabili di giunto è importante essere in grado di specificare la posizione dei link rispetto agli altri della catena. Lo scopo della *cinematica diretta* è quello di determinare la posizione e l'orientamento dell'end effector in funzione delle variabili di giunto. Ad ogni link i si associa un sistema di riferimento (x_i, y_i, z_i) , abbreviato con *sdr*, usando, per esempio, il formalismo introdotto dalla rappresentazione di Denavit-Hartenberg (DH) [17]. La relazione fra il sdr $i-1$ e il sdr i è data dalla matrice di trasformazione

$$A_i = \begin{bmatrix} \cos \theta_i & -\cos \alpha_i \sin \theta_i & \sin \alpha_i \sin \theta_i & a_i \cos \theta_i \\ \sin \theta_i & \cos \alpha_i \cos \theta_i & -\sin \alpha_i \cos \theta_i & a_i \sin \theta_i \\ 0 & \sin \alpha_i & \cos \alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (2.11)$$

I parametri del link i α_i , angolo di *twist*, e a_i , lunghezza del link, sono fissati e resi noti nelle specifiche del costruttore. Le variabili di giunto θ_i e d_i , sono fisse o variabili in base al tipo di giunto, rotoidale o prismatico. Di conseguenza la matrice A_i è in funzione di un solo parametro variabile, θ_i o d_i , mentre tutti gli altri valori sono fissati per uno specifico giunto (Fig. 2.9). Se un manipolatore ha n link, il vettore q delle variabili di giunto risulterà essere un vettore di dimensione n contenente una combinazione di θ e d . La matrice A è una matrice di trasformazione omogenea nella forma

$$A_i = \begin{bmatrix} R_i & p_i \\ 0 & 1 \end{bmatrix} \quad (2.12)$$

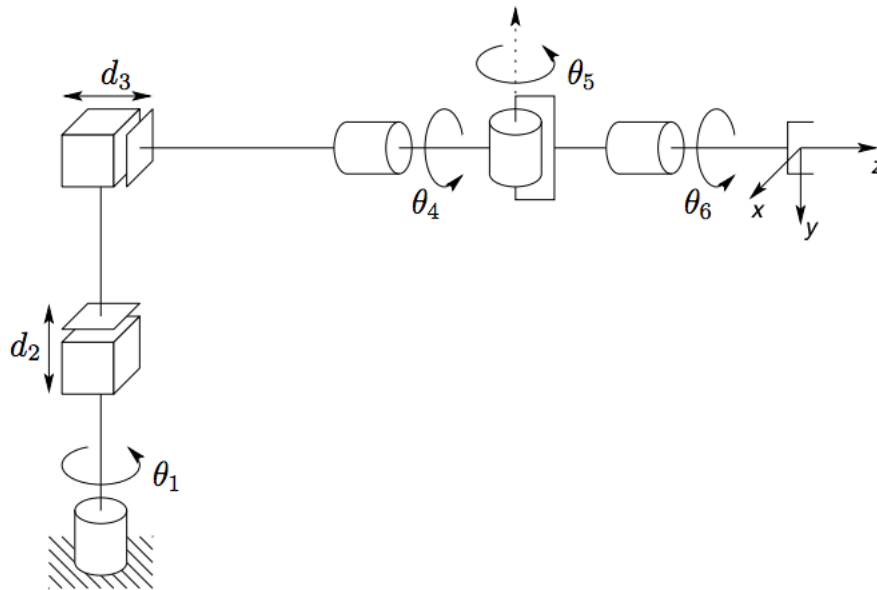


Figura 2.9: Esempio di manipolatore in cui sono stati evidenziati la tipologia di giunti e le relative variabili.

dove R_i è la matrice di rotazione e p_i è il vettore di traslazione. In questo modo se r_i è un punto descritto rispetto al sdr del link i , lo stesso punto ha coordinate r_{i-1} rispetto al sdr del link $i-1$ date da

$$r_{i-1} = A_i r_i \quad (2.13)$$

La *trasformazione omogenea* è una matrice 4×4 che può descrivere sia le rotazioni che le traslazioni, di conseguenza il vettore che descrive la posizione in un dato sdr è di dimensione 4 e della forma

$$r_i = \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} \quad (2.14)$$

con (x_i, y_i, z_i) coordinate del punto nel sdr i . Nello specifico è possibile scrivere la relazione (2.17) esplicitando la rotazione e traslazione come

segue

$$\begin{bmatrix} x_{i-1} \\ y_{i-1} \\ z_{i-1} \end{bmatrix} = R_i \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} + p_i \quad (2.15)$$

ponendo il sdr $i-1$ come fisso e il sdr i come ruotato e traslato. A_i può essere quindi vista come la trasformazione della rappresentazione del vettore r_i nel sdr i nella rappresentazione r_{i-1} nel sdr $i-1$. D'altra parte è anche la descrizione del sdr i in termini del sdr $i-1$, descrivendo l'orientamento degli assi tramite la rotazione R_i e la posizione dell'origine del sdr i nel sdr $i-1$ tramite il vettore p_i .

Per ottenere le coordinate di un punto nel sdr 0 (o base) si può usare la matrice

$$T_i = A_1 A_2 \dots A_i. \quad (2.16)$$

Date le coordinate r_i di un punto espresso nel sdr del link i le coordinate dello stesso punto nel sdr base saranno quindi

$$r_0 = T_i r_i. \quad (2.17)$$

T_i è detta *catena cinematica di trasformazioni*. Si può definire la matrice T del braccio come

$$T = T_n = A_1 A_2 \dots A_n \quad (2.18)$$

con n il numero di link del manipolatore. Ogni oggetto descritto in termini del sdr n -simo può di conseguenza essere espresso in coordinate del sdr base. La posizione dell'end effector è data, in coordinate del sdr base, dall'ultima colonna della matrice T .

La *cinematica inversa* consiste nel determinare la configurazione che il robot dovrebbe adottare per ogni link per raggiungere la posizione di un obiettivo con un determinato orientamento tramite l'end effector. La risoluzione di questo problema è di notevole importanza per poter trasformare

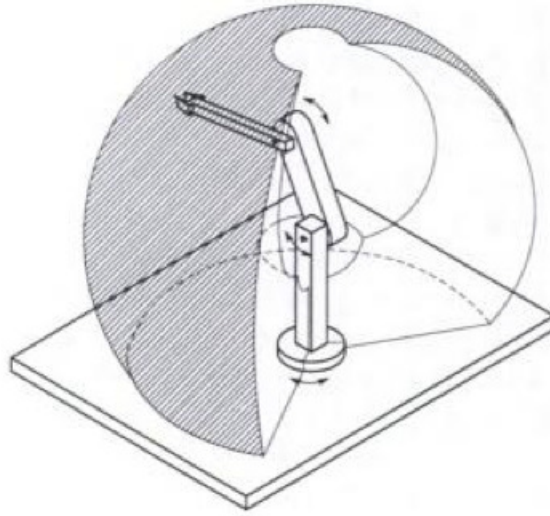


Figura 2.10: Manipolatore sferico con relativo spazio di lavoro.

le specifiche del moto, assegnate all'end effector nello spazio di lavoro, nel corrispondente moto dei singoli giunti per permettere di eseguire il movimento desiderato. Solitamente è molto più difficile rispetto ad un problema di cinematica diretta ed è necessario sviluppare tecniche efficienti e specifiche per la particolare struttura cinematica di ogni manipolatore.

Mentre il problema di cinematica diretta ha sempre un'unica soluzione che può essere ottenuta semplicemente valutando le equazioni di cinematica, il problema della cinematica inversa può o non può avere soluzioni. Anche se una soluzione esiste, può o non può essere unica. Inoltre essendo le equazioni, in generale, funzioni non lineari nelle variabili di giunto, le soluzioni sono difficili da ottenere anche quando esistono.

Spazio di lavoro e ridondanza

Lo *spazio di lavoro* è quel volume descritto dall'origine dell'end effector quando tutti i giunti del manipolatore eseguono tutti i movimenti possibili. Lo spazio di lavoro dipende strettamente dalla geometria del manipolatore e dai vincoli meccanici dei giunti (Fig. 2.10).

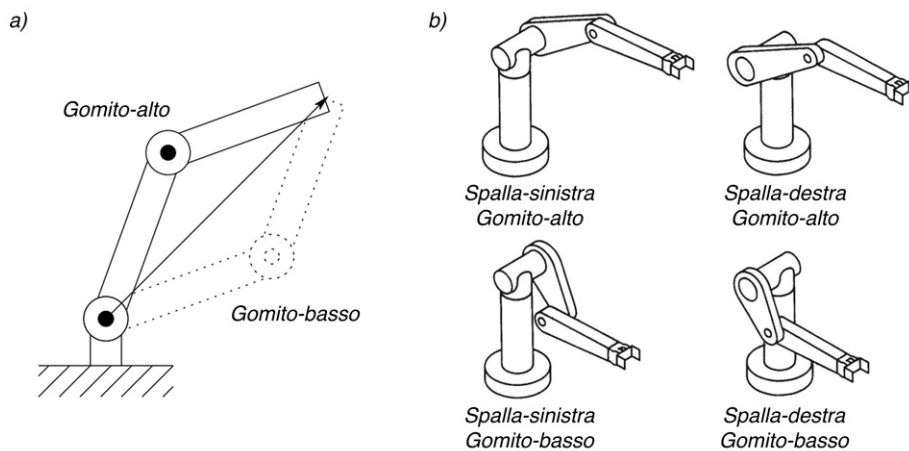


Figura 2.11: Esempio di ridondanza nel manipolatore RR (a) e nel manipolatore RRR (b).

Un manipolatore è definito *ridondante* quando ha un numero di gradi di libertà maggiore del numero di variabili strettamente necessarie per portare a termine un compito. La ridondanza implica che un compito a livello end effector possa essere eseguito in diversi modi a livello dei giunti (esistono infinite soluzioni alla cinematica inversa). Un manipolatore RR planare permette di ottenere una data posizione dell'end effector tramite due distinte configurazioni dei giunti, chiamate gomito-alto e gomito-basso (Fig. 2.11a). Un manipolatore RRR può raggiungere un punto nello spazio tridimensionale con 4 configurazioni: spalla-destra gomito-alto, spalla-destra gomito-basso, spalla-sinistra gomito-alto, spalla-sinistra gomito-basso (Fig. 2.11b) [18].

La ridondanza è un concetto relativo al compito assegnato al manipolatore; un manipolatore può essere ridondante rispetto a un compito e non esserlo rispetto a un altro. Per esempio, un robot planare con tre giunti è ridondante rispetto al posizionamento nel piano 2D, ma non rispetto al posizionamento e orientamento nel piano.

Perché quindi utilizzare intenzionalmente un braccio ridondante? La scelta è dettata dal fatto che si vuole dotare il manipolatore di destrezza e ver-

satilità nel moto. Il classico esempio è costituito dal braccio umano che ha sette gradi di mobilità: tre nella spalla, uno nel gomito e tre nel polso, senza considerare i gradi di mobilità delle dita. Questo manipolatore è intrinsecamente ridondante infatti permette di muovere, ad esempio, il gomito mantenendo ferma la posizione della mano e della spalla, tutto ciò a vantaggio della mobilità e della velocità. Così, per esempio, è possibile evitare ostacoli nello spazio di lavoro. Inoltre, se un giunto di un manipolatore raggiunge il suo limite meccanico, ci potrebbero essere altri giunti che permettono comunque l'esecuzione del moto dell'end effector che si era deciso. La ridondanza può essere usata anche per evitare singolarità cinematiche, minimizzare i consumi di energia, ottimizzare il tempo di percorrenza e aumentare l'affidabilità del sistema.

2.4 Apprendimento

2.4.1 Developmental Learning

La ricerca sullo sviluppo umano e la ricerca sulla costruzione di robot intelligenti può e dovrebbe essere complementare. Studi sullo sviluppo umano hanno prodotto una grossa varietà di teorie, modelli e costrutti sperimentali che sono stati a lungo una fonte di ispirazione per sistemi robotici. Non solo la robotica può ispirarsi allo sviluppo umano ma anche lo sviluppo umano può trarre profitto dall'opportunità di valutazione e sperimentazione che la robotica offre; studiare le fasi dello sviluppo è infatti estremamente difficile a causa della complessa interazione che intercorre fra il soggetto, i meccanismi di apprendimento e le dinamiche ambientali [6], [7].

Naturale *pro* artificiale: come la psicologia dello sviluppo impatta sulla robotica

La psicologia dello sviluppo è spesso utilizzata dalla robotica come fonte di ispirazione. I problemi che vengono affrontati nella psicologia dello sviluppo (come la modalità con cui i neonati imparano ad orientarsi verso gli stimoli e come imparano a muoversi in luoghi sconosciuti) sono di interesse anche della comunità robotica. Modelli della psicologia dello sviluppo spesso offrono decomposizioni comportamentali e osservazioni su prestazioni che possono fornire uno schema per architetture software. Inoltre, gli schemi di sviluppo incrementale sono stati adottati come tecniche di valutazione per i sistemi robotici.

Lo sviluppo umano sfrutta un incremento graduale sia nella complessità interna (percettiva e motoria) sia nella complessità esterna (compiti e complessità ambientali regolati dall'istruttore) per ottimizzare l'apprendimento di nuove abilità. Questo tipo di apprendimento è detto *developmental learning*. Approcci simili applicati alla robotica permettono di costruire sistemi che sono in grado di scalare su compiti e problemi più complessi ottimizzando l'apprendimento in una maniera simile a quello che succede nel neonato. Possono essere appresi nuovi comportamenti più sofisticati sfruttando un incremento graduale in complessità sia interna che esterna e riutilizzando strutture ed informazioni acquisite da comportamenti imparati precedentemente.

Naturale *via* artificiale: come la robotica può contribuire alla psicologia dello sviluppo

La ricerca sulla robotica umanoide può contribuire anche ad indagare in merito a questioni scientifiche riguardanti la natura dell'intelligenza umana. I robot umanoidi possono rivelarsi degli strumenti unici per la ricerca nelle scienze cognitive. L'implementazione robotica di modelli cognitivi,

comportamentali e di sviluppo forniscono un banco di prova per valutare la potenza predicibile e la validità di tali modelli. Un modello implementato in un robot permette di effettuare dei test molto più accurati e in secondo luogo validarli tramite degli esperimenti controllati e ripetibili. Lievi modifiche agli esperimenti possono essere usate per isolare e valutare singoli fattori (sia ambientali che interni) indipendentemente dalle tante variabili che disturbano le normali osservazioni sui comportamenti. Gli esperimenti possono anche essere ripetuti con condizioni quasi identiche al fine di permettere una facile validazione. Inoltre, le strutture interne dei modelli possono essere manipolate per osservare gli effetti quantitativi e qualitativi dei comportamenti. Un modello robotico può anche essere sottoposto ad un test controverso che è potenzialmente pericoloso, costoso o immorale se condotto su un umano; le "condizioni di contorno" del modello possono essere esplorate testando diversi addestramenti e condizioni ambientali. Infine, un modello robotico può essere utilizzato per suggerire e valutare una potenziale strategia di intervento prima di essere applicata ai soggetti umani.

Sviluppo infantile e motor babbling

I primi anni di vita rappresentano il momento in cui avvengono il maggior numero di cambiamenti nelle capacità cognitive e sensomotorie di un uomo. In questo periodo le abilità motorie, percettive e cognitive maturano, mentre fisicamente il corpo cambia in maniera drastica. Le scienze che si occupano dello studio sullo sviluppo umano sembrano concordare sull'importanza dell'attività motoria e sul suo ruolo nella percezione [3]. Questa correlazione non dovrebbe sorprenderci se si pensa che molte proprietà dell'ambiente sono percepite in maniera diretta solo tramite l'interazione fisica con l'ambiente stesso. La forma di un oggetto, la consistenza o il tipo di superficie sono caratteristiche che vengono percepite meglio tramite

l'uso del tatto. L'attività motoria ha pertanto un ruolo fondamentale per il corretto sviluppo percettivo dell'uomo. Un'ipotesi è che sviluppo percettivo e motorio progrediscano di pari passo.

Alcuni studiosi ritengono che l'abilità di percepire visivamente le caratteristiche degli oggetti quali ad esempio volume, durezza e consistenza, non emerga prima dei 6-9 mesi, mentre la capacità di distinguere forme tridimensionali si sviluppi solamente verso i 12-15 mesi di vita. Tale tempistica corrisponde a quella dello sviluppo motorio, l'esplorazione degli oggetti e la conseguente abilità di percepirne determinate proprietà richiede infatti un elevato grado di controllo del movimento. Nei bambini, durante lo sviluppo, l'impossibilità di eseguire una certa azione potrebbe infatti impedire la percezione della corrispondente proprietà.

Nei primi mesi di vita le capacità motorie dei neonati sono limitate. La coordinazione e il controllo dei movimenti delle braccia risultano poco puliti, solo raramente e in maniera quasi casuale il bambino riesce a toccare gli oggetti e ad afferrarli. Intorno ai tre mesi queste abilità migliorano e i movimenti di prensione (*grasping*) diventano più affidabili, anche se l'oggetto è quasi sempre afferrato con la mano aperta sfruttando l'opposizione tra le dita e il palmo (*power grasp*). Questa interazione con gli oggetti pur essendo semplice è comunque sufficiente a rendere note proprietà come temperatura, dimensione/volume e durezza. Infine, verso i 6-9 mesi, i bambini iniziano a manipolare gli oggetti con una certa abilità e con differenti strategie che includono l'uso delle dita in differenti configurazioni. A questo punto dello sviluppo, grazie al maggior controllo sugli oggetti, i bambini iniziano a percepire altre proprietà più complesse quali la forma e il tipo di superficie. Intorno ai nove mesi d'età, la capacità di interagire con gli oggetti in maniera ripetitiva e accurata consente di maturare ulteriormente le capacità cognitive. Le azioni di prensione sugli oggetti permettono di costruire una rappresentazione multisensoriale che combina

opportunamente esperienza visiva, tattile e chinestetica.

I neonati mostrano movimenti spontanei delle braccia (*motor babbling*, termine coniato da Meltzoff e Moore [8]) già nell'utero. Von Hofsten [9] suggerisce che questi movimenti giocano un ruolo fondamentale nello sviluppo dei muscoli, delle articolazioni e della struttura del sistema nervoso centrale. In altre parole l'importanza di queste attività motorie autoprodotte, che possono essere considerate come azioni esplorative, trova giustificazione nel mapping di base fra differenti parti del sistema nervoso. Come von Hofsten fa notare, le azioni esplorative tradizionalmente sono sempre state pensate come azioni focalizzate sul mondo esterno e sugli oggetti ed eventi in esso contenuti, ma le azioni esplorative possono anche essere focalizzate sul proprio sistema d'azione. Piaget [10] enfatizza che quando i neonati compiono movimenti ripetutamente, essi stanno in effetti esplorando il proprio sistema d'azione. L'ipotesi di reazione-circolare di Piaget è uno dei più influenti principi sulla teoria dello sviluppo motorio nei bambini. I bambini imparano la coordinazione occhio-mano sulla base della produzione di movimenti spesso casuali che aumentano l'associazione tra motori efferenti e percezione re-afferente. Questa associazione è successivamente usata come base su cui costruire un comportamento sensomotorio più complesso.

Per un sistema robotico complesso autonomo, come un robot umanoide, l'apprendimento basato sul *motor babbling* è considerato un metodo efficace per sviluppare autonomamente un modello interno del proprio corpo e dell'ambiente. Uno dei principali problemi nell'apprendimento è che il dominio di apprendimento è troppo grande per essere completamente coperto (*frame problem*). Una strategia efficace consiste nell'aumentare la velocità di apprendimento mantenendo alta la qualità; una strategia basata sul campionamento casuale è spesso considerato l'approccio più robusto per un dominio sconosciuto.

L'obiettivo del motor babbling è di campionare lo spazio di tutte le possibili configurazioni dei giunti in un modo il più possibile esaustivo e imparare come associare le azioni con la percezione sensoriale. Questo metodo è molto dipendente dall'hardware usato; per esempio, alcune configurazioni possono fisicamente danneggiare il robot. Di conseguenza, alcuni vincoli potrebbero essere necessari quando si effettua il campionamento su tutte le possibili configurazioni. In generale lo scopo è di campionare sempre più nel dettaglio per capire come differenti configurazioni influiscono sui sensori. Inoltre, per un addestramento più robusto, ogni movimento ha bisogno di essere eseguito diverse volte, causando un numero di campioni di dati sensoriali per ogni possibile movimento. Questo è necessario perché i sensori sono sensibili al rumore e all'ambiente.

Un esempio di applicazione robotica: coordinazione mano-occhio

Diamond [4] ha mostrato che i neonati fra i 5 e i 12 mesi progrediscono attraverso un numero distinto di fasi nello sviluppo del reaching guidato dalla vista. In questa progressione, i neonati nelle fasi finali dimostrano delle sofisticate strategie di reaching per recuperare un giocattolo in scenari più impegnativi. Quando l'abilità di reaching del neonato è stata sviluppata, i passaggi successivi servono per migliorare la competenza appresa precedentemente.

In robotica, senza una prospettiva di sviluppo incrementale, il problema di indicare un obiettivo visivo è un problema di mapping sensomotorio $\mathbb{R}^2 \rightarrow \mathbb{R}^6$: la posizione dell'obiettivo in coordinate visive (bidimensionali) devono essere convertite in una traiettoria per i 6 gradi di libertà del braccio. Usando la decomposizione comportamentale di Diamond osservata nei neonati, Marjanovic [5] ridusse questa funzione $\mathbb{R}^2 \rightarrow \mathbb{R}^6$ in una coppia di funzioni apprese $\mathbb{R}^2 \rightarrow \mathbb{R}^2$ e una funzione fissa non degenerata $\mathbb{R}^2 \rightarrow \mathbb{R}^4$ con espliciti segnali di errore. Da una prospettiva esterna, il com-

portamento del robot è abbastanza rudimentale. Dato uno stimolo visivo, tipicamente un ricercatore che muove un oggetto davanti alle camere, il robot esegue un movimento saccadico per foveare il target e poi raggiunge con le braccia il target. I primi movimenti sono inaccurati e spesso nelle direzioni sbagliate, ma dopo poche ore di pratica l'accuratezza migliora drasticamente. Per raggiungere un target visivo, il robot deve imparare a mappare dalle coordinate visive del target $x = (x, y)$ alle coordinate dei giunti del braccio $a = (a_0, \dots, a_5)$. Per ottenere questo, il robot prima deve imparare a foveare il target usando una mappa di saccadi $S : x \mapsto e$ che mette in relazione le posizioni nel piano immagine della camera con i comandi motori (angoli) necessari a foveare l'occhio in quella posizione $e = (\text{pan}, \text{tilt})$. Una volta che il target è foveato, il robot deve imparare un movimento balistico mappando coordinate testa-centriche e , in coordinate braccio-centriche a . Per semplificare il problema relativo al controllo dei 6 gradi di libertà del braccio, le posizioni del braccio sono specificate come una combinazione lineare di primitive di base di postura precedentemente apprese.

2.4.2 Reti neurali RBF

Per mantenere basso il costo computazionale del calcolo del mapping visomotorio, attenersi ad un approccio di matrice biologica e semplificare il calcolo della cinematica inversa del braccio robotico, si è scelto di avvicinarsi al problema con l'utilizzo di reti neurali. Le reti neurali RBF (radial basis function) sono un caso particolare di modello lineare che permette di ipotizzare o stimare una funzione, a partire da delle coppie di input e output fornite, senza avere un'approfondita conoscenza della funzione che le lega (approccio supervisionato) [14]. I due vantaggi principali delle reti neurali sono: mantenere i calcoli matematici semplici (sono problemi di algebra lineare) e il costo computazionale relativamente basso. Inoltre es-

se si adattano facilmente a molteplici contesti ed eventuali variazioni del sistema.

Nell'apprendimento supervisionato, la funzione è appresa dagli esempi forniti da un "insegnante". Il set di addestramento contiene coppie di variabile indipendente (input) e variabile dipendente (output), tipicamente sono dei vettori. La funzione prende quindi la forma

$$f(x) = \sum_{j=1}^m w_j h_j(x) \quad (2.19)$$

il modello $f(x)$ è espresso come combinazione lineare di un set di m funzioni fissate chiamate solitamente funzioni base (basis functions) [13]. La flessibilità di $f(x)$, la sua capacità di approssimare differenti funzioni, deriva solo dalla libertà di scegliere diversi valori per i pesi. Le funzioni base, ed ogni parametro che possono contenere, sono fisse; se esse potessero variare durante il processo di apprendimento, il modello non sarebbe lineare. Come basis function può essere usata qualsiasi famiglia di funzioni, una classe di speciale interesse è quella delle funzioni a base radiale. La loro caratteristica principale è che la loro risposta cresce (o decresce) monotonicamente con la distanza da un punto centrale. Il centro, la scala di distanza, e la forma precisa della funzione, sono tutti parametri del modello (fissi se lineare). Una tipica funzione a base radiale è la gaussiana, nel caso di input scalare é:

$$h(x) = \exp\left(-\frac{(x-c)^2}{r^2}\right) \quad (2.20)$$

con c parametro del centro e r del raggio (detto anche spread). Le RBF con funzioni gaussiane sono locali (danno una risposta significativa solo in prossimità del centro) e sono anche biologicamente plausibili [15], in quanto il comportamento di ogni singola funzione di base è paragonabile a quello di un neurone, contrariamente rispetto alle funzioni multiquadratiche che, simili a una gaussiana rovesciata, danno una risposta in tutto lo spazio. Inoltre, le reti neurali RBF si basano sul fatto che qualsiasi funzione

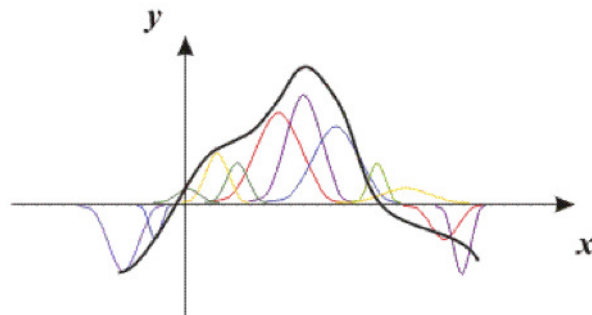


Figura 2.12: Approssimazione con funzioni gaussiane.

continua in un intervallo può essere approssimata con la precisione desiderata da una somma di funzioni gaussiane scelte appropriatamente, si veda Fig. 2.12.

Sebbene le reti neurali RBF possano avere un qualsiasi numero di livelli nascosti (*hidden layers*), tradizionalmente se ne usa uno soltanto (Fig. 2.13). Le singole funzioni di base dell'hidden layer possono essere definite neuroni, convenzionalmente il numero di neuroni dello strato nascosto è pari al numero dei dati di training. Questo può portare però ad overfitting e richiede parecchio tempo per addestrare i vari parametri. Il concetto di overfitting è molto importante nell'apprendimento automatico. Solitamente un algoritmo di apprendimento viene allenato usando un certo insieme di esempi. Si assume che l'algoritmo di apprendimento raggiungerà uno stato in cui sarà in grado di predire gli output per tutti gli altri esempi che ancora non ha visionato, cioè si assume che il modello di apprendimento sarà in grado di *generalizzare*. Tuttavia, soprattutto nei casi in cui l'apprendimento è stato effettuato troppo a lungo, il modello potrebbe adattarsi a caratteristiche che sono specifiche solo del training set, ma che non hanno riscontro nel resto dei casi; perciò, in presenza di overfitting, le prestazioni (cioè la capacità di adattarsi o prevedere) sui dati di allenamento aumenteranno, mentre le prestazioni sui dati non visionati saranno peggiori. È anche vero, d'altro canto, che una rete neurale con troppi pochi neuroni

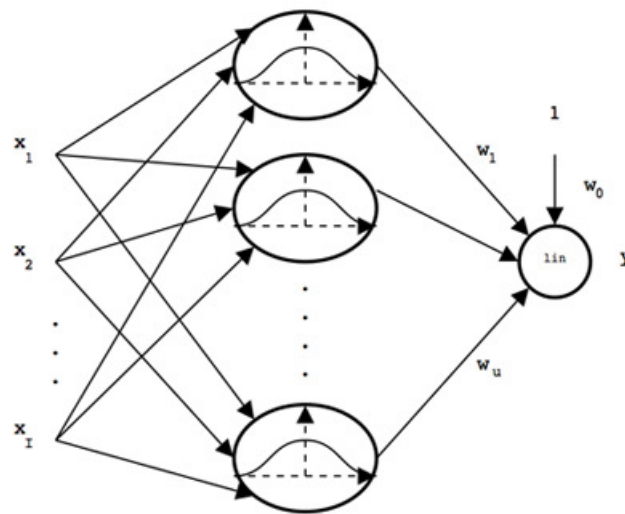


Figura 2.13: Rappresentazione rete neurale RBF con un solo hidden layer di funzioni gaussiane.

non è in grado di modellizzare la funzione e porta ad un basso livello di accuratezza. Tendenzialmente si seleziona il numero adeguato di neuroni tramite un processo sperimentale iterativo.

Ogni neurone (funzione gaussiana di base) è definito dalla posizione della sua media, dalla covarianza e dal coefficiente di combinazione lineare con le altre gaussiane, detto peso. Ogni funzione copre solo una piccola regione dello spazio. Volendo ricostruire la funzione che definisce i dati in ingresso e uscita su tutto lo spazio, è necessario distribuire su tutto lo spazio delle feature un numero adeguato di centri. Risulta evidente che, ogni rete neurale RBF è caratterizzata da un grande numero di parametri che devono essere determinati dal processo di training; fortunatamente esistono molte tecniche euristiche che permettono di definirli senza procedere iterativamente con tentativi per ognuno di essi. Il risultato non è ottimo in maniera assoluta, ma la rete funziona comunque ragionevolmente bene.

Capitolo 3

Lavori correlati: visione e reaching

Nonostante il problema della rappresentazione del *body schema* sia ampiamente trattato in letteratura (vedi [22] per una review), per l'obiettivo del nostro lavoro consideriamo solo i paper nei quali ci si approccia al mapping visomotorio attraverso il motor babbling [23]. Nel campo della visione sono già state sviluppate numerose applicazioni, riassunte in [24]. Gli autori descrivono le problematiche che sorgono ad esempio, nel riconoscimento di oggetti, nel tracking, nella manipolazione, localizzazione e mapping. Sono state proposte molte strategie per gestire il controllo a basso livello di un sistema di stereovisione attivo, molti approcci propongono l'utilizzo di reti neurali basate sul modello di disparità dell'energia per controllare la foveazione. Wang ad esempio, illustra lo sviluppo autonomo del controllo della vergenza massimizzando la risposta neurale attraverso il metodo dell'apprendimento per rinforzo [25] [26]. Gibaldi modella un sistema che estrae direttamente la disparità-vergenza senza un calcolo esplicito della disparità [27]. Inoltre, lo stesso autore implementa una strategia di controllo per la testa iCub permettendole di foveare oggetti fissi o mobili lungo la direzione della profondità considerando solo alcune configurazioni fis-

se nella direzione di tilting [28]. Shimonura propone una testa hardware costruita in FPGA (Field Programmable Gate Array) con retine in silicene; il sistema di vergenza fovea un punto calcolando la disparità tramite il modello energetico [29]. Saegusa propone un metodo per inferire il body schema basato su un motore stocastico di babbling, guidato dall'esperienza visomotoria [30].

Per quanto riguarda l'apprendimento supervisionato del compito di raggiungimento di un oggetto, si veda la ricerca proposta da Burnod e collaboratori [31]. Questo modello sfrutta la struttura a gradiente dell'architettura parieto-frontale sinergicamente con le proprietà combinatorie dei neuroni. In particolare propone che il mappaggio sensomotorio sia compiuto in maniera progressiva dal graduale coinvolgimento di *matching units* appartenenti a domini combinatori diversi. L'apprendimento sinaptico è il risultato del rafforzarsi delle connessioni fra due unità dovuto alla loro ripetuta co-attivazione. Di seguito vengono descritti i quattro passaggi che, secondo la loro ricerca, sono necessari ad allineare la posizione della mano (in coordinate dei giunti) con una posizione obiettivo (in coordinate visive). Ogni passaggio corrisponde all'apprendimento in diversi gruppi di *matching units*.

Motor babbling: il soggetto muove l'end effector (i.e. mano) in maniera casuale. Vengono quindi co-attivate le unità che inviano ai muscoli il segnale motorio e quelle che ricevono il segnale somato-motorio propriocettivo. Da quest'azione risulta un rafforzamento del legame fra questa coppia di set di unità che, dopo l'apprendimento, permetterà di generare il segnale adeguato da inviare ai muscoli per ottenere in risposta i segnali sensoriali attesi.

Controllo dello sguardo: quando lo sguardo si muove verso uno stimolo si ha l'attivazione delle unità che inviano il segnale di movimento ai muscoli degli occhi e quelle della retina che vengono stimulate dalla visione. Questo permette di focalizzare l'attenzione in una determinata area dello spazio.

Foveazione della mano: quando gli occhi guardano la mano che si muove l'input che segnala il movimento dei muscoli degli occhi e quello dei muscoli della mano vengono confrontati. Si rinforzano in questo modo le connessioni che mettono in relazione *pathway* equivalenti di mano e occhi imparando quindi a seguire i movimenti della mano stessa.

Raggiungimento di obiettivi foveati: quando la mano si muove verso un obiettivo foveato, la co-attivazione delle connessioni mano-sguardo e sguardo-retina precedentemente rinforzate porta al rafforzamento delle connessioni delle unità che permetteranno di raggiungere col movimento della mano una posizione obiettivo, anche se questa non si trova nel campo visivo.

Piaget propone un paradigma riguardante l'apprendimento della catena cinematica inversa attraverso l'uso del protocollo di apprendimento a reazione circolare [10]. Il concetto della reazione circolare è descritto da Piaget con il seguente loop comportamentale: un bambino compie movimenti casuali, generati spontaneamente col suo braccio, i suoi occhi seguono il movimento del braccio permettendo di imparare una trasformazione dalla rappresentazione visuale del braccio ad una rappresentazione dello stesso nel mondo dei giunti. Inspirati da questo processo, i ricercatori hanno implementato un processo adattativo simile per apprendere la cinematica. Questo processo è basato su un ciclo comportamentale controllato autonomamente diviso in due fasi: produzione e percezione, è chiamato motor babbling [8].

Durante la fase di motor babbling, il sistema di controllo genera endogenamente dei movimenti casuali che posizionano il braccio in diverse posizioni con la mano all'interno del campo visivo. Per ogni posizione del braccio, il sistema di visione identifica la mano ed apprende la correlazione fra variabili dei giunti e degli occhi. Dopo che le correlazioni sono salvate

come pesi delle connessioni, il segnale visivo in ingresso passa attraverso la rete pesata e produce una predizione dei vettori di attivazione necessari da inviare al braccio. La differenza fra i vettori predetti e quelli generati esternamente è usata per modificare i pesi della rete. Dopo numerose iterazioni, la rete potrà stimare con precisione come raggiungere i punti foveati.

Sulla stessa linea di pensiero, ma con un'attenzione particolare alla disposizione topologica dei neuroni all'interno della rete neurale (lattice neurale 3D), si può annoverare Martinez ed i suoi collaboratori [32].

Uno studio più esteso e recente riguardante la ricostruzione della cinematica inversa del robot Hoap3 con 24 gradi di libertà è quello di Hersch, Sauser e Billard [33]. Viene ricostruito il body schema tramite l'osservazione dei movimenti delle estremità degli arti, in seguito a variazioni della posizione dei giunti intermedi. Si mette in evidenza come il tracking della posizione dei giunti intermedi possa portare ad una più rapida convergenza del sistema.

Implementazione di una modellizzazione dell'area V6A

A titolo d'esempio riguardo i risultati finora raggiunti, si riporta il recente (2011) lavoro di Chinellato [21] che implementa su un robot umanoide una modellizzazione semplificata dell'area V6A. Occhi e braccia possono essere considerati come effettori separati che ricevono i controlli dei motori da una mappa sensomotoria implicita condivisa dello spazio peripersonale. Il risultato dei loro movimenti è utilizzato per aggiornare la mappa stessa. Chinellato propone di codificare implicitamente la mappa sensomotoria dello spazio peripersonale in una struttura di funzioni a base radiale, esse modellano le funzioni dell'area V6A e le sue connessioni con aree puramente visuali e premotrici. Si può supporre che l'area V6A contenga neuroni che hanno solo risposta visiva, neuroni apparentemente coinvolti prevalentemente nell'azione di movimento e altri neuroni attivati

in tutte le fasi del processo, dediti a garantire un'efficace trasformazione da un sistema di riferimento all'altro.

Dapprima si apprende la trasformazione fra spazio visuale e spazio dei giunti dei motori di braccio e occhi, implementando delle reti neurali RBF ispirate ai neuroni dell'area V6A. Chinellato in questo stadio, per semplificare, utilizza solo un braccio a due giunti ed elimina dagli occhi il movimento di tilting in modo che l'ambiente accessibile sia posizionato su un piano orizzontale 2D di fronte al soggetto. I differenti tipi di neuroni V6A possono essere simulati con popolazioni di funzioni a base radiale con campo ricettivo distribuito secondo diversi criteri. In una prima approssimazione, Chinellato, distribuisce uniformemente nello spazio visivo alcuni neuroni che assumeranno il ruolo di quelli visuali che si erano prevalentemente attivati durante le fasi di fix e hold; i neuroni motori, corrispondenti alla fase move, sono organizzati secondo lo spazio dei giunti; i neuroni intermedi possono essere posizionati secondo diversi criteri (spazio visuale, dei giunti o di versione e vergenza) in modo da codificare il sistema di riferimento posizionato nella testa del robot.

Fissando e raggiungendo una posizione, sia liberamente che verso una direzione specifica, l'agente apprende come eseguire la cinematica diretta ed inversa fra stereovisione, coordinate oculomotorie e rappresentazione dello spazio dei giunti. La mappa sensomotoria integrata che permette di rappresentare contestualmente lo spazio peripersonale attraverso differenti visioni e parametri motori non è mai costruita esplicitamente, ma emerge grazie all'interazione con l'ambiente. Le principali informazioni sensoriali usate per costruire questa rappresentazione sono retino-ottiche e propriocettive. Contestualmente i motori degli occhi e del braccio sono gestiti separatamente, questo permette di mantenere entrambi fissati sull'obiettivo e stabilire una rappresentazione comune dello spazio entro cui si trova. Il framework di consapevolezza spaziale proposto è implementato quindi a

partire da un codice comune, ottenuto da una procedura di apprendimento basata sull'analisi degli errori nel raggiungimento dell'obiettivo. Chinellato definisce quest'approccio come *self supervised learning*, nel quale le differenti modalità si supervisionano a vicenda, i movimenti degli occhi e del braccio migliorano ed ottengono una rappresentazione man mano più precisa dello spazio. Considerando che per codificare n differenti segnali, servirebbero tanti neuroni quanti i possibili valori dei segnali elevati alla potenza di n , risulta evidente come la situazione risulterebbe ingestibile anche per il cervello. È per questo motivo che risulta più logico mantenere una rappresentazione del mondo corpo-centrica alla quale i sistemi sensomotori e di visione accedono per ottenere e aggiornare le informazioni. Questo approccio risulta inoltre coerente con il funzionamento dell'area V6A del cervello, infatti come abbiamo visto in precedenza, in essa è identificabile una ripartizione dei neuroni secondo lo stesso principio: alcuni codificano la posizione in coordinate retinocentriche, altri rispetto all'arto e altri ancora in entrambi i modi, in modo da poter trasformare i sistemi di riferimento. È quindi una scelta coerente mantenere una struttura modulare che separa il processo di trasformazione da coordinate retiniche a corpo-centriche e dal corpo a centrate nell'arto.

Per quanto riguarda la rappresentazione dell'informazione binoculare, Chinellato assume che il problema della corrispondenza dell'immagine acquisita da due fotocamere sia risolto ed ipotizza l'acquisizione di un'immagine ciclopica abbinata all'informazione di disparità. Per mantenere una fedeltà biologica le telecamere non verranno mai mosse separatamente ed in maniera incoerente, il movimento è quindi gestito tramite angoli di versione e vergenza invece che con una posizione assoluta. Ulteriori semplificazioni introdotte da Chinellato riguardano il collo, che essendo mantenuto fisso, porta il sistema di riferimento del corpo a corrispondere con quello centrato nella testa; inoltre elimina il feedback tattile dal braccio meccanico

ed il controllo dell'arto è basato unicamente sulla propriocezione.

Trasformazione da coordinate manuali a oculomotorie

Apprendere la trasformazione da immagine binoculare a posizione degli occhi consiste nell'identificare un obiettivo e fovearlo con entrambi gli occhi, in modo da associare alle sue posizioni movimenti appropriati degli angoli di versione e vergenza. Come input visuale viene usato un campo visivo ciclopico accompagnato dall'informazione di disparità e la trasformazione viene implementata con una rete RBF per le ragioni illustrate in precedenza. Le migliori performance vengono ottenute (nel caso di Chinellato) con unità di forma gaussiana per entrambi gli input, descritte dalla seguente funzione vettoriale equivalente ad una gaussiana a singola variabile:

$$h_i(x) = e^{-(x-c_i)^T \Sigma_i^{-1}(x-c_i)} \quad (3.1)$$

$$u_k = \sum_{i=1}^{n_k} w_{i,k} h_i(x) \quad (3.2)$$

Il processo di apprendimento consiste nel trovare i pesi che adattano meglio i dataset d'ingresso con le uscite. La posizione dei centri viene stabilita in modo fisso nonostante con una disposizione variabile si sarebbero potute ottenere migliori performance; viene effettuata questa scelta in modo da favorire una migliore plausibilità biologica. È per questo motivo che le funzioni vengono distribuite ispirandosi al criterio retino-ottico seguendo una distribuzione logaritmica dei centri, infatti in questo modo si simula la maggior concentrazione di sensori visivi nell'area della fovea e si ha una codifica più raffinata per le piccole disparità. Nello studio a cui facciamo riferimento viene stabilita una soglia d'errore di 0,10 mm per $p=1000$ punti di training, ottenuta utilizzando 49 neuroni in un lattice neurale 7×7 , distribuiti come in Fig. 3.1.

Per ottenere i dataset si esegue la simulazione seguente:

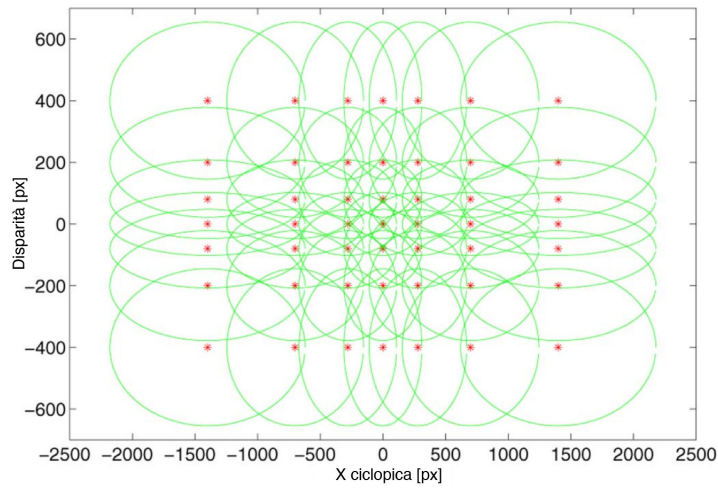


Figura 3.1: Distribuzione e spread delle funzioni a base radiale sullo spazio ciclopico/disparità, nella rete neurale per il calcolo delle coordinate oculomotorie a partire da quelle visuali.

1. L'agente fissa un punto dato nello spazio, vicino a uno stimolo visuale.
2. Esegue un movimento degli occhi casuale verso una posizione differente.

La coppia di dati in ingresso ed uscita è quindi composta dalla posizione visuale (coordinata ciclopica e disparità) dello stimolo iniziale visto dalla posizione corrente degli occhi e dalla versione e vergenza del movimento eseguito. I pesi vengono settati inizialmente con la tecnica di batch learning tipica delle RBF:

$$w_0 = (H^T * H)^{-1} * H^T * Y \quad (3.3)$$

Dopo uno studio sull'evoluzione dell'errore, si determina che il numero adeguato di punti del dataset sia 200. Per simulare una sorta di apprendimento da parte del robot, viene eseguito un training passo passo per i successivi 200 punti; questa seconda fase è eseguita applicando la discesa del gradiente con *delta rule* per aggiornare i pesi ad ogni *timestep*:

$$w_{t+1} = w_t - \alpha * h_t^T * (y_t - u_t) \quad (3.4)$$

L'utilizzo di questa tecnica permette di mantenere il sistema flessibile a possibili cambiamenti dell'accuratezza visiva e della cinematica del corpo.

Trasformazione da coordinate oculomotorie a braccio

La seguente fase è divisa in due stage: movimenti liberi e movimenti orientati a un obiettivo. La fase a movimenti liberi consiste nel far compiere al robot dei movimenti casuali per poi muovere gli occhi in modo da fissare la posizione finale della mano; in questo modo si apprende una trasformazione da mondo dei giunti a quello delle coordinate oculomotorie e viceversa. La seconda fase deve necessariamente seguire la prima, in quanto necessita della trasformazione per permettere agli occhi di fissare un punto nello spazio e muovere poi il braccio in quella direzione. Anche in questo caso il posizionamento automatico dei centri guidato dai punti in ingresso sarebbe la soluzione standard, viene però favorita la plausibilità biologica. Per focalizzare l'attenzione sulla distribuzione dei neuroni, Chinellato semplifica il braccio robotico rendendo accessibile ad esso solo uno spazio 2D posizionato orizzontalmente di fronte agli occhi del soggetto e procede testando distribuzioni omogenee sia nello spazio dei giunti che in quello di versione e vergenza, con funzioni gaussiane e sigmoidali (Fig. 3.2). La soluzione migliore e biologicamente coerente si configura essere una disposizione dei neuroni mista: 50 neuroni, 25 distribuiti in modo omogeneo nello spazio dei giunti e 25 nello spazio di versione e vergenza (Fig. 3.3). L'errore è dell'ordine di 5mm ed è paragonabile a quello ottenuto con la distribuzione automatica dei centri, inoltre la rete neurale mista richiama la presenza nell'area V6A di diversi tipi di neuroni. Esperimenti più recenti dimostrano che il campo ricettivo di molti neuroni è distribuito sull'area di vergenza e versione, facendo passare in secondo piano l'importanza della distribuzione nello spazio dei giunti.

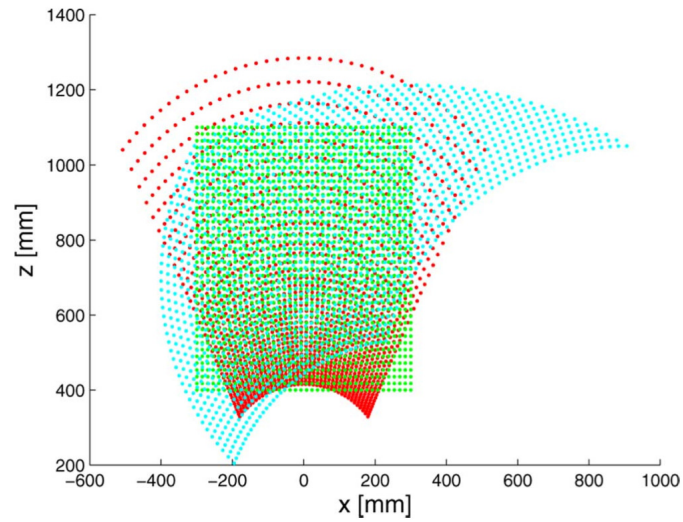


Figura 3.2: Mapping dello spazio sovrapponendo in rosso lo spazio oculomotorio con una distribuzione uniforme di vergenza/versione, in azzurro lo spazio dei giunti del braccio RR ed in verde lo spazio cartesiano.

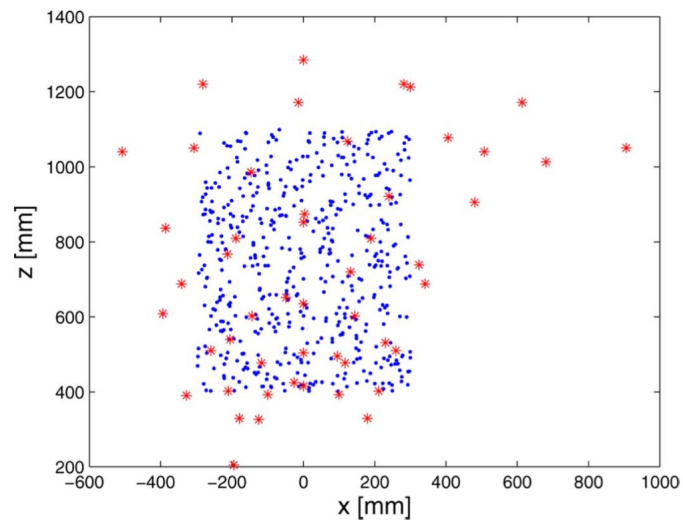


Figura 3.3: In rosso distribuzione delle funzioni a base radiale secondo un criterio misto (25 oculomotori + 25 nello spazio del braccio), in blu invece è visualizzato lo spazio dei punti di addestramento in coordinate cartesiane.

Capitolo 4

Sistema

Nella prima parte di questo capitolo illustreremo come sono costituiti il sistema di visione con collo fisso (base) e quello con l'aggiunta della rotazione nel collo (sistema esteso). Nella seconda parte sono descritte le strutture dei bracci robotici RR e RRR che abbiamo scelto di usare per simulare la parte di reaching.

4.1 Visione

4.1.1 Sistema base

Alla luce dei capitoli precedenti illustriamo ora come è costituito il sistema di stereovisione bio-ispirato sotteso alle reti neurali che generalmente implementano la foveazione. Le equazioni fondamentali del controllore sono basate sulla legge di uguale innervazione di Hering (trattata nella Sezione 2.2.6), che permette di ottenere un controllo preciso delle traiettorie senza imporre dei vincoli specifici sugli angoli del sistema di visione.

Il sistema è un modello proporzionale che deve essere allenato per apprendere i parametri caratteristici. Il controllore è strutturato con tre gradi di libertà, due per il comando di panning per entrambi gli occhi ed uno per il tilting. Il modello sviluppato tratta la versione, vergenza e tilting come tre

movimenti indipendenti e li considera in parallelo. È importante osservare che, per estrarre informazioni rilevanti riguardo la disparità, l'obiettivo deve essere foveato il più rapidamente possibile, in quanto una stima affidabile della disparità è fattibile solo nei dintorni della fovea. Il controllore dei giunti di versione, vergenza e tilting è quindi rappresentato dalle seguenti equazioni:

$$\dot{\theta}_{\text{versione}} = K_1(x_L + x_R) \quad (4.1)$$

$$\dot{\theta}_{\text{vergenza}} = K_2\delta \quad (4.2)$$

$$\dot{\theta}_{\text{tilt}} = K_3(y_L + y_R) \quad (4.3)$$

dove (x_L, y_L) e (x_R, y_R) sono rispettivamente le coordinate del punto target proiettato nel piano immagine sinistro e destro; δ è la disparità fra le due proiezioni del target quando le disparità verticali sono trascurabili $y_L \approx y_R$; K_1 , K_2 e K_3 sono costanti proporzionali che determinano quanto velocemente i movimenti si adattano alla legge di Hering. Considerando la prima equazione, la versione può essere inizializzata anche da uno stimolo che raggiunge un solo occhio, ricadendo quindi nel caso in cui l'obiettivo si trova all'estremità sinistra o destra del campo visivo, e non sia quindi visibile ad entrambi gli occhi. Tuttavia, la vergenza non è possibile perché la versione si annulla quando l'obiettivo cade nel punto uguale e opposto nella fovea sinistra e destra. Quindi, il controllo della vergenza diventa possibile solo attraverso la seconda equazione, dove la disparità gioca un ruolo fondamentale.

Gli angoli di panning θ_l e θ_r della telecamera sinistra e della telecamera destra e l'angolo comune di tilting θ_t possono essere ottenuti con:

$$\dot{\theta}_r = \dot{\theta}_{\text{versione}} - \dot{\theta}_{\text{vergenza}} \quad (4.4)$$

$$\dot{\theta}_l = \dot{\theta}_{\text{versione}} + \dot{\theta}_{\text{vergenza}} \quad (4.5)$$

$$\dot{\theta}_t = -\dot{\theta}_{\text{tilt}} \quad (4.6)$$

La catena di trasformazioni che porta dal sistema di riferimento del mondo a quello delle telecamere é:

$$\mathbf{R}_W^{L/R} = \mathbf{R}_H^{L/R}(\theta_{L/R}) \mathbf{R}_W^H(\theta_T) \quad (4.7)$$

dove $\mathbf{R}_W^{L/R}$ è la rototraslazione fra il sistema di riferimento del mondo ed il sistema di riferimento della telecamera (sinistra o destra), $\mathbf{R}_H^{L/R}(\theta_{L/R})$ è la rototraslazione fra il sistema di riferimento della testa e quello della telecamera e $\mathbf{R}_W^H(\theta_T)$ è il comando di tilting definito come una rotazione del sistema di riferimento testa rispetto al sistema del mondo.

Modello delle telecamere

Per mantenere l'approccio il più possibile human-like, non solo per quanto riguarda l'approccio metodologico e qualitativo, ma anche quantitativo, si è scelto di simulare il sistema di visione con parametri simili a quelli medi umani. La distanza fra il centro degli occhi è di 0.06 metri, lo spazio extra-personale di fronte al soggetto è profondo 1.4 metri e largo 2.8 metri. Lo spazio extra-personale è definito come la porzione di spazio che si estende oltre la sfera peripersonale di manipolazione ed è fondamentale per l'orientamento [34]. Il campo visivo del singolo occhio è di 140 gradi, da cui è possibile ricavare la lunghezza focale f applicando un semplice calcolo geometrico: $\alpha = \beta/2$, $\tan(\alpha) = 320/f$ come mostrato in Fig. 4.1.

Questi parametri hanno condotto alla realizzazione della matrice di calibrazione delle telecamere (4.8), uguale per entrambe in modo da essere il più consistenti possibili con la realtà. A riprova dei dati ipotizzati in precedenza, una lunghezza focale di 120 pixel con un piano immagine di 640x480 pixel conducono ad un angolo della lente di 140 gradi. Le matrici sono *undistorte e non rettificate*, in quanto si suppone di rapportarsi con un sistema

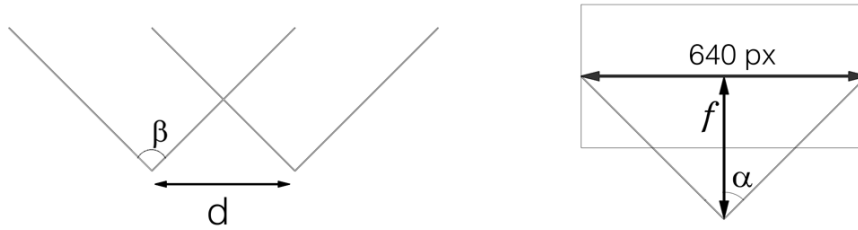


Figura 4.1: A sinistra rappresentazione del sistema occhi umano: la distanza d interpupillare è di 0.06m, il campo visivo β di un singolo occhio di 140° . A destra la rappresentazione di una singola telecamera della quale si deve calcolare la focal length f .

attivo e si considera il modello delle telecamere consistente.

$$\begin{bmatrix} 120 & 0 & 320 \\ 0 & 120 & 240 \\ 0 & 0 & 1 \end{bmatrix} \quad (4.8)$$

L'origine del sistema di riferimento del collo coincide con l'origine del sistema di riferimento del mondo; per ora l'unico movimento consentito al collo è il tilting e le telecamere sono quindi traslate di 0.2 metri lungo l'asse Y (Fig. 4.2).

4.1.2 Sistema esteso

Il modello sviluppato sinora, prende in considerazione un sistema con soli tre gradi di libertà per foveare una generica posizione nello spazio 3D visibile. Si estende quindi il modello aggiungendo un ulteriore grado di libertà nel collo per migliorare le prestazioni della testa nell'attività di panning.

Allo scopo di aggiungere nel collo un'ulteriore giunto di rotazione, si sono ipotizzate diverse versioni del controllore. Abbiamo introdotto la componente per il collo Eq. (4.9), aggiungendola alle regole (4.1), (4.2) e (4.3)

$$\dot{\theta}_{\text{collo}} = K_4(x_L + x_R) \quad (4.9)$$

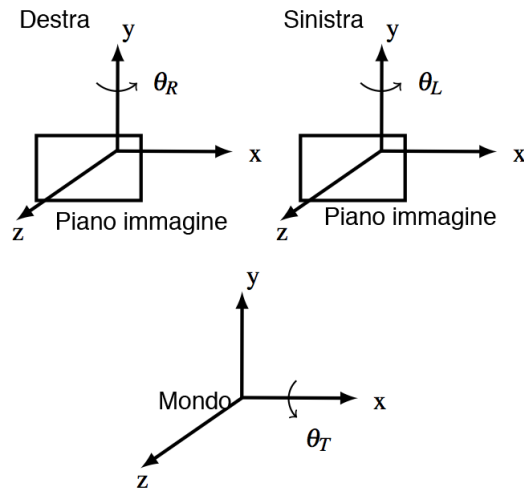


Figura 4.2: Sistemi di riferimento del sistema base. Il movimento di tilting è eseguito lungo l'asse x del sistema mondo e ruota entrambi i sistemi di riferimento degli occhi di θ_T [rad]. Idealmente, abbiamo definito un collo virtuale che esegue il movimento di tilting.

Questo comporta che i movimenti del collo dipendano dalla posizione del punto da foveare nel piano immagine. I movimenti del collo consistono solamente in una rotazione lungo l'asse Y e sono indipendenti dal comando di tilting. Introducendo un nuovo grado di libertà per il collo si è reso il sistema ridondante, si rende necessario ridefinire come segue la catena di trasformazioni che porta dal sistema di riferimento del mondo a quello delle telecamere:

$$R_W^{L/R} = R_N^{L/R}(\theta_{L/R}) R_H^N(\theta_N) R_W^H(\theta_T) \quad (4.10)$$

dove $R_W^{L/R}$ è la rototraslazione fra il sistema di riferimento del mondo ed il sistema di riferimento della telecamera (sinistra o destra), $R_N^{L/R}(\theta_{L/R})$ è la rototraslazione fra il sistema di riferimento del collo e quello della telecamera, $R_H^N(\theta_N)$ è la rototraslazione fra la testa ed il collo (definita come movimento lungo la direzione di panning) e $R_W^H(\theta_T)$ è il comando di tilting definito come una rotazione del sistema di riferimento testa rispetto al sistema del mondo. Per calcolare gli angoli dei movimenti di pan, tilt e

rotazione, le equazioni

$$\dot{\theta}_{\text{versione}} = K_1(x_L + x_R) \quad (4.11)$$

$$\dot{\theta}_{\text{vergenza}} = K_2\delta \quad (4.12)$$

$$\dot{\theta}_{\text{tilt}} = K_3(y_L + y_R) \quad (4.13)$$

$$\dot{\theta}_{\text{collo}} = K_4(x_L + x_R) \quad (4.14)$$

devono essere combinate in modo opportuno: chiamiamo configurazioni i diversi modi in cui si possono ottenere i movimenti. Le configurazioni sono riassunte nella Tabella 4.1 e riflettono i seguenti concetti:

- il movimento degli occhi (panning) può essere mediato dal movimento del collo (configurazioni 1-4);
- il movimento del collo (panning) può essere mediato da vergenza e versione (configurazioni 1,2,3,6);
- il movimento di occhi e collo possono essere indipendenti l'uno dall'altro (configurazione 5).

4.2 Reaching

Abbiamo intrapreso lo studio del sistema di reaching in due fasi distinte, affrontando il problema in maniera incrementale: dapprima si è definita la struttura di un braccio robotico planare RR di ispirazione umana, poi abbiamo provato ad aggiungere al braccio un ulteriore grado di libertà in corrispondenza della spalla per permettere all'end effector di raggiungere qualsiasi punto dello spazio 3D e di conseguenza avvicinarci ulteriormente a una struttura fisica umanoide.

<i>Configurazione 1</i>	<i>Configurazione 2</i>
$\dot{\theta}_R = \dot{\theta}_{\text{versione}} - \dot{\theta}_{\text{vergenza}} + \dot{\theta}_{\text{collo}}$ $\dot{\theta}_L = \dot{\theta}_{\text{versione}} + \dot{\theta}_{\text{vergenza}} + \dot{\theta}_{\text{collo}}$ $\dot{\theta}_T = -\dot{\theta}_{\text{tilt}}$ $\dot{\theta}_N = \dot{\theta}_R$	$\dot{\theta}_R = \dot{\theta}_{\text{versione}} - \dot{\theta}_{\text{vergenza}} + \dot{\theta}_{\text{collo}}$ $\dot{\theta}_L = \dot{\theta}_{\text{versione}} + \dot{\theta}_{\text{vergenza}} + \dot{\theta}_{\text{collo}}$ $\dot{\theta}_T = -\dot{\theta}_{\text{tilt}}$ $\dot{\theta}_N = \dot{\theta}_L$
<i>Configurazione 3</i>	<i>Configurazione 4</i>
$\dot{\theta}_R = \dot{\theta}_{\text{versione}} - \dot{\theta}_{\text{vergenza}} + \dot{\theta}_{\text{collo}}$ $\dot{\theta}_L = \dot{\theta}_{\text{versione}} + \dot{\theta}_{\text{vergenza}} + \dot{\theta}_{\text{collo}}$ $\dot{\theta}_T = -\dot{\theta}_{\text{tilt}}$ $\dot{\theta}_N = \dot{\theta}_{\text{collo}} - \dot{\theta}_{\text{versione}}$	$\dot{\theta}_R = \dot{\theta}_{\text{versione}} - \dot{\theta}_{\text{vergenza}} + \dot{\theta}_{\text{collo}}$ $\dot{\theta}_L = \dot{\theta}_{\text{versione}} + \dot{\theta}_{\text{vergenza}} + \dot{\theta}_{\text{collo}}$ $\dot{\theta}_T = -\dot{\theta}_{\text{tilt}}$ $\dot{\theta}_N = \dot{\theta}_{\text{collo}}$
<i>Configurazione 5</i>	<i>Configurazione 6</i>
$\dot{\theta}_R = \dot{\theta}_{\text{versione}} - \dot{\theta}_{\text{vergenza}}$ $\dot{\theta}_L = \dot{\theta}_{\text{versione}} + \dot{\theta}_{\text{vergenza}}$ $\dot{\theta}_T = -\dot{\theta}_{\text{tilt}}$ $\dot{\theta}_N = \dot{\theta}_{\text{collo}}$	$\dot{\theta}_R = \dot{\theta}_{\text{versione}} - \dot{\theta}_{\text{vergenza}}$ $\dot{\theta}_L = \dot{\theta}_{\text{versione}} + \dot{\theta}_{\text{vergenza}}$ $\dot{\theta}_T = -\dot{\theta}_{\text{tilt}}$ $\dot{\theta}_N = \dot{\theta}_{\text{collo}} - \dot{\theta}_{\text{versione}}$

Tabella 4.1: Configurazioni provate per il sistema di visione esteso.

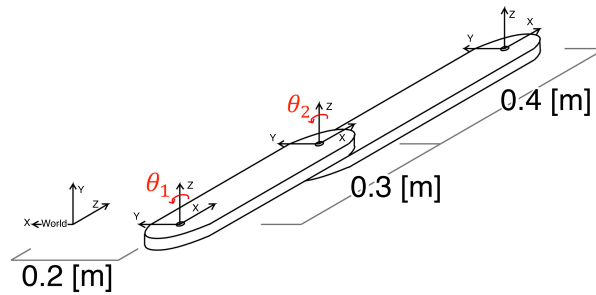


Figura 4.3: Struttura del braccio robotico RR. θ_1 e θ_2 sono le due variabili dei giunti di rotazione.

4.2.1 Braccio RR

Abbiamo implementato per prima cosa il braccio planare a due giunti di rotazione RR, per renderlo il più simile possibile ad un braccio umano (si è scelto il braccio destro) l'abbiamo posizionato spostato a destra rispetto alla base del collo (coincidente con l'origine del mondo) di 20 centimetri e quindi anche di 20 centimetri verso il basso rispetto agli occhi. La lunghezza dell'arto è di 70 centimetri divisi in 40 centimetri di avambraccio e 30 di braccio (Fig. 4.3).

Con questa struttura, i punti dello spazio accessibili sono quelli posizionati di fronte al soggetto, sul piano XZ del sistema mondo. Il braccio potrà raggiungerli con due diverse configurazioni e quindi in modo ridondante: una con il gomito rivolto verso destra e l'altra verso sinistra. La posizione dell'end effector può essere calcolata nel sistema di riferimento tramite la seguente catena di rototraslazioni:

$$R_H^W = R_A^W R_F^A(\theta_1) R_H^F(\theta_2) \quad (4.15)$$

dove R_H^W è la rototraslazione fra il sistema di riferimento della mano ed il sistema di riferimento mondo, R_A^W è la rototraslazione fra il sistema di riferimento del braccio e quello del mondo, $R_F^A(\theta_1)$ è la rototraslazione fra l'avambraccio e il braccio (definita come rotazione sull'asse Z del giunto rotoidale) e $R_H^F(\theta_2)$ è la rotazione del sistema di riferimento della mano rispet-

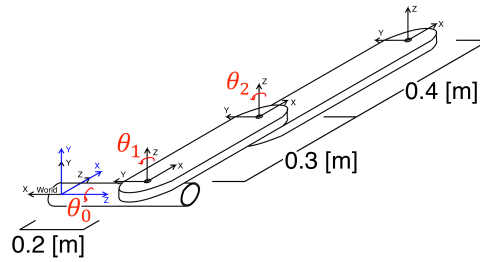


Figura 4.4: Struttura del braccio robotico RRR. θ_0 , θ_1 e θ_2 sono le tre variabili dei giunti di rotazione.

to al sistema dell'avambraccio. Le coordinate dell'end effector nel sistema di riferimento mondo si possono ricavare dall'ultima colonna di R_H^W .

4.2.2 Braccio RRR

Abbiamo pensato di procedere verso un obiettivo più sfidante ed implementare un braccio a tre giunti di rotazione. Per mantenere un approccio human oriented abbiamo scelto di posizionare l'ulteriore giunto di rotazione in corrispondenza dell'origine del sistema di riferimento mondo simulando una spalla di 20 centimetri, collegato alla spalla un avambraccio di 30 centimetri e ad esso il braccio di 40 centimetri (Fig. 4.4).

Lo spazio raggiunto dal braccio a tre giunti di rotazione passa dall'essere planare a tridimensionale, si assiste quindi all'introduzione di una nuova variabile, essa rappresenta il terzo giunto del braccio. Si rende necessario ridefinire come segue la catena di trasformazioni che porta dal sistema di riferimento della mano a quello del mondo:

$$R_H^W = R_S^W R_A^S(\theta_0) R_F^A(\theta_1) R_H^F(\theta_2) \quad (4.16)$$

dove R_H^W è la rototraslazione fra il sistema di riferimento della mano ed il sistema di riferimento mondo, R_S^W è la rototraslazione fra il sistema di riferimento della spalla e quello del mondo, $R_A^S(\theta_0)$ è la rototraslazione fra il sistema di riferimento del braccio e quello della spalla (con θ_0 nuova varia-

bile di giunto che permette l'elevamento del braccio), $R_F^A(\theta_1)$ è la rototraslazione fra l'avambraccio e il braccio (definita come rotazione sull'asse Z del giunto rotoidale) e $R_H^F(\theta_2)$ è la rotazione del sistema di riferimento della mano rispetto al sistema dell'avambraccio. Le coordinate dell'end effector nel sistema di riferimento mondo si possono ricavare dall'ultima colonna di R_H^W .

Questo tipo di configurazione introduce un'ulteriore ridondanza per il raggiungimento di un punto nello spazio infatti ogni punto è raggiungibile da 4 configurazioni diverse del braccio.

Capitolo 5

Metodi

In questo capitolo vengono esposti i metodi utilizzati per affrontare l'implementazione del nostro sistema visomotorio. Nella prima parte illustreremo come è stato possibile ricavare i parametri proporzionali del controllore attraverso la minimizzazione dell'errore nel piano immagine. Nella seconda parte verranno introdotti i vincoli per i bracci robotici e i metodi necessari per addestrare e convalidare la rete neurale usata per il reaching.

5.1 Visione

5.1.1 Sistema base

Per valutare le performance del sistema di visione usiamo la seguente stima dell'errore:

$$e_{L/R} = \sqrt{x_{L/R}^2 + y_{L/R}^2} \quad (5.1)$$

che è la distanza euclidea calcolata sul piano immagine fra la posizione dell'obiettivo da foveare ed il centro del piano immagine che modella la fovea (in questo caso abbiamo l'origine del sistema di riferimento del piano immagine esattamente al centro del piano immagine stesso). I pedici L/R si riferiscono rispettivamente alla telecamera sinistra e destra. Si è scelto di

valutare l'errore separatamente per l'occhio sinistro e destro per capire se l'errore di foveazione varia fra i due occhi.

Per garantire un errore minimo per ogni punto 3D che deve essere foveato, indipendentemente dalle posizioni iniziali delle due telecamere, è necessario apprendere i parametri K applicando la seguente tecnica di minimizzazione:

$$c(X, Y, Z) = e_L^2 + e_R^2 + \sum_j |\dot{\theta}_L|_j + \sum_j |\dot{\theta}_R|_j + \sum_j |\dot{\theta}_T|_j \quad (5.2)$$

$$K = \operatorname{argmin}_{K_1, K_2, K_3} \sum_x \sum_y \sum_z c(x, y, z)$$

Le distanze Euclidee sono necessarie per valutare le performance del sistema nel foveare i punti desiderati, le sommatorie minimizzano la lunghezza delle traiettorie percorse e quindi evitano oscillazioni intorno alla posizione finale. La funzione obiettivo è minimizzata numericamente utilizzando il metodo della discesa gradiente. I punti utilizzati per l'addestramento coprono in maniera omogenea il campo visivo, possono essere descritti nel seguente modo:

$$x \in [x_{\min}, x_{\max}]$$

$$y \in [y_{\min}, y_{\max}]$$

$$z \in [z_{\min}, z_{\max}]$$

Addestriamo il robot con punti nello spazio extrapersonale, quindi i range sono:

$$x \in [-1.4, 1.4] \text{ metri} \quad \text{con step di 0.7 metri}$$

$$y \in [-1.4, 1.4] \text{ metri} \quad \text{con step di 0.7 metri}$$

$$z \in [0.2, 1.4] \text{ metri} \quad \text{con step di 0.4 metri}$$

Lo spazio viene campionato in 100 punti ma il dataset utilizzato per l'addestramento è costituito solo dai 71 punti visibili da entrambi gli occhi;

si è scelto uno step abbastanza grande per avere pochi punti ed evitare overfitting.

5.1.2 Sistema esteso

Successivamente si è poi adattata la procedura di apprendimento utilizzata per il sistema a tre gradi di libertà in modo che potesse gestire un sistema a quattro gradi di libertà:

$$c(X, Y, Z) = e_L^2 + e_R^2 + \sum_j |\dot{\theta}_{Lj}| + \sum_j |\dot{\theta}_{Rj}| + \sum_j |\dot{\theta}_{Tj}| + \sum_j |\dot{\theta}_{Nj}| \quad (5.3)$$

Il sottospazio usato per l'addestramento è analogo a quello usato nel sistema base.

5.2 Reaching

5.2.1 Braccio RR

La struttura del braccio RR permette di raggiungere ogni punto del piano con due diverse configurazioni e quindi in modo ridondante: una con il gomito rivolto verso destra e l'altra verso sinistra. L'unica plausibile per il braccio destro di un uomo è la prima, quindi imponiamo come vincolo che le escursioni massime dei giunti di rotazione siano di $[-\frac{\pi}{2}, \frac{\pi}{2}]$ per l'angolo della spalla θ_1 e di $[0, \frac{3}{4}\pi]$ per l'angolo del gomito θ_2 , in modo che sia sempre mantenuta una postura naturale dell'arto nel generare il dataset di giunti per addestrare la rete neurale. Lo spazio di lavoro ora raggiungibile è mostrato in Fig. 5.1

Generazione dataset

Ci proponiamo di simulare il processo d'apprendimento di un bambino e quindi ispirandoci al developmental learning implementare il motor

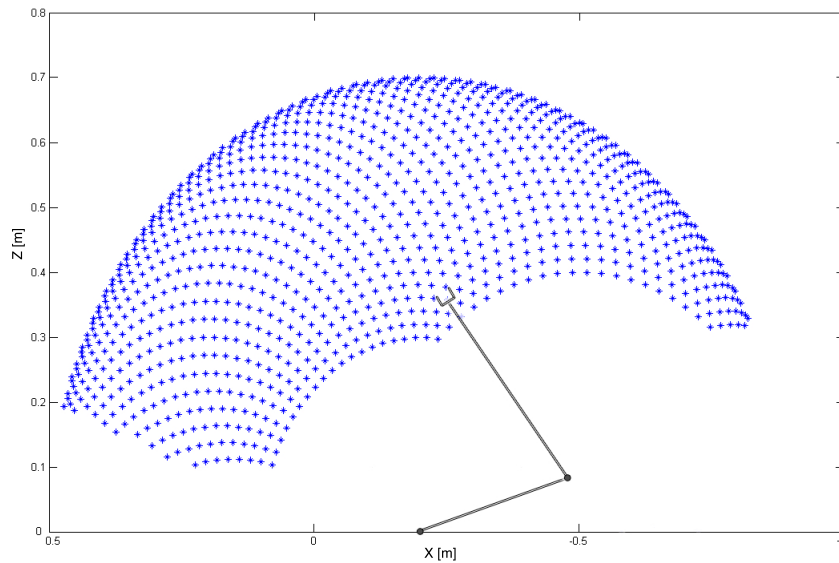


Figura 5.1: Spazio di lavoro del braccio robotico RR. I vincoli introdotti negli angoli di giunto fanno sì che non si ottenga un semicerchio pieno come spazio di lavoro.

babbling (si veda Sezione 2.4.1). Come un bambino nelle prime fasi dell'apprendimento spaziale muove le braccia in maniera pseudocasuale senza un obiettivo specifico, il nostro robot simulato porterà il braccio nei diversi punti dello spazio 3D. Questo movimento viene compiuto senza la volontà specifica di raggiungere un certo obiettivo nello spazio, ma suddividendo i range di movimento dei giunti in diversi passi e testando tutte le possibili combinazioni di posizione dei giunti. In questo modo si porta l'end effector del braccio a muoversi in posizioni distribuite uniformemente su tutto lo spazio di lavoro che rientra nel campo visivo delle telecamere in situazione di riposo.

Per apprendere un legame fra la direzione dello sguardo e il movimento del braccio necessario per raggiungere il punto di fissazione, anche i bambini, inizialmente, seguono con gli occhi la direzione verso la quale si sta muovendo la loro mano. Anche il nostro robot simula un comportamento

simile: sono in corso diverse ricerche, nel campo della visione, con l'obiettivo di permettere ad una macchina di percepire e rilevare stimoli interessanti nel campo visivo e portare lo sguardo in quella direzione; supponiamo ora di definire l'estremità del braccio come "stimolo interessante", non è difficile immaginare l'implementazione di un sistema in grado di osservare la propria mano che si muove casualmente nello spazio dei giunti. Si rammenti che non abbiamo a disposizione un sistema di gestione dell'attenzione in grado di scindere la posizione dell'end effector (stimolo interessante) dalle altre informazioni contenute nelle immagini delle due telecamere e che il nostro sistema è simulato. Per ottenere le coordinate motorie in cui gli occhi si sposterebbero per osservare la mano, a partire dalla configurazione dei giunti del braccio e tramite la cinematica diretta, abbiamo ottenuto la posizione dell'end effector nello spazio 3D; da tale punto, conoscendo le relative proiezioni nei piani immagine del sistema stereovisivo delle telecamere, si è passati alle coordinate motorie di occhi e collo ottenute grazie alla realizzazione del controllore proporzionale che abbiamo descritto in maniera estesa nella prima parte della Sezione 4.1.2.

In questo modo abbiamo potuto abbinare coordinate dei giunti del braccio con coordinate motorie del sistema visivo per mantenere lo sguardo posizionato sulla mano. Il passo successivo consiste nello svolgere il compito inverso e quindi implementare il sistema di trasformazione dello spazio di riferimento gestito dall'area del cervello V6A. Ciò significa, partendo dalla direzione dello sguardo, calcolare in maniera istantanea grazie alle reti neurali che descriveremo fra poco, le coordinate in cui i giunti del braccio si devono portare per raggiungere la feature.

Rete neurale

La rete neurale integrata nel nostro sistema ha l'obiettivo di fornire, a fronte di un segnale in ingresso costituito dagli angoli motori degli occhi,

un'uscita corrispondente agli angoli dei giunti del braccio per raggiungere il punto foveato. Per implementarla ci si è avvalsi della funzione di matlab *newrb*, che crea ed addestra la rete neurale grazie all'algoritmo di minimizzazione dell'MSE (Mean Squared Error). Per un buon funzionamento della rete uno dei problemi principali consiste nello scegliere la coppia di valori spread e soglia che garantiscono il miglior compromesso fra precisione e capacità di generalizzare. Una volta creato il dataset di punti per l'addestramento si è proceduto testando sperimentalmente diversi valori di spread e goal. Si è osservato che generalmente un numero di centri e quindi neuroni pari a 100, sono sufficienti per ottenere buoni risultati. Oltre i 100 neuroni e fino a 200 le performance della rete non peggiorano, ma l'aumento di precisione non è giustificato considerando l'aumento di tempo necessario per il training. Con un numero di neuroni superiore a 200 assistiamo a fenomeni di overfitting: le prestazioni della rete peggiorano, i risultati diventano meno omogenei e compaiono dei valori con un errore notevolmente maggiore dell'errore medio. Sono state realizzate diverse reti neurali, ognuna con i valori dei parametri più adatti a modellare i differenti tipi di braccio robotico che sono stati utilizzati, verranno descritte mano a mano che verranno trattati i diversi bracci.

Avendo a disposizione le coppie di dati con cui addestrare la rete neurale suddivisi in

- input: angoli del sistema di visione per foveare la mano,
- target: angoli dei giunti del braccio,

é possibile passare al training. Grazie alla fase di addestramento si crea una rete neurale RBF che ricevendo in ingresso gli angoli del sistema di visione delle telecamere che stanno foveando, fornisce in uscita gli angoli che i giunti del braccio devono raggiungere. La plausibilità biologica e la sem-

plicità di design delle reti RBF hanno spinto la nostra scelta verso questa direzione.

Si è scelto di addestrare la rete neurale e quindi posizionare i centri tramite un processo iterativo basato sulla minimizzazione dell'errore medio (MSE).

1. Si scelgono uno spread per l'ampiezza della gaussiana dei centri ed una soglia per l'errore massimo tollerabile.
2. Fra i dati di input si seleziona una tupla di valori ed in quella posizione viene fissato un centro.
3. Si addestra la rete.
4. Si simula la rete neurale e si verifica dove si trova l'errore massimo.
5. Nella posizione dell'errore massimo viene posizionato un nuovo centro.
6. Si ripete a partire dal terzo punto fino al raggiungimento della soglia fissata.

La scelta dello spread e della soglia influenzano il numero di neuroni che verrà utilizzato per approssimare la rete. Un numero troppo alto di neuroni può portare a rallentamenti nell'addestramento e problemi di overfitting che conducono la rete alla perdita della sua capacità di generalizzare e fornire output plausibili per dati in input non presenti nel dataset di training.

Il problema principale nell'addestramento della rete neurale si riduce di conseguenza ad essere la scelta di un'adeguata dimensione per lo spread delle funzioni di base che rappresentano i neuroni ed una soglia di tolleranza per l'errore; il numero ed il posizionamento dei centri deriva da questi

valori. Per stabilire quali sono i parametri che minimizzano l'errore abbiamo testato un range di valori (sia di spread che di precisione) su ogni set di dati generati dalla k-fold Cross-validation.

La Cross-validation è un metodo statistico che permette di confrontare e validare modelli derivati tramite algoritmi di apprendimento da set di dati. Si dividono i dati in due porzioni: uno utilizzato per addestrare un modello e l'altro utilizzato per validarlo. Si implementa la k-fold Cross-validation per verificare la robustezza della rete neurale appresa e la sua capacità di produrre risultati affetti da poco errore nel momento in cui ricevono in ingresso nuovi valori su cui non è stata addestrata. Nella k-fold validation

- si dividono i dati in k segmenti della stessa dimensione detti: *fold*.
- Sono eseguite k iterazioni di addestramento utilizzando $k - 1$ fold.
- Per ogni ciclo di addestramento si esclude un fold, esso verrà utilizzato per la validazione.

In questo modo, ogni valore viene utilizzato sia per il training che per il testing e si verifica la capacità della rete neurale di generalizzare a fronte di valori sconosciuti.

Per illustrare come abbiamo implementato questo metodo ci avvaliamo della Fig. 5.2, descriviamo le varie fasi:

1. Il rettangolo bianco rappresenta l'intero dataset dei valori in input alla rete neurale, cioè gli angoli dei giunti del sistema visivo e del braccio necessari per foveare un punto e raggiungerlo. Il rettangolo blu sono i punti nello spazio 3D corrispondenti agli angoli.
2. I valori degli angoli dei giunti sono stati calcolati discretizzando ordinatamente lo spazio dei giunti del braccio, calcolando la cinematica

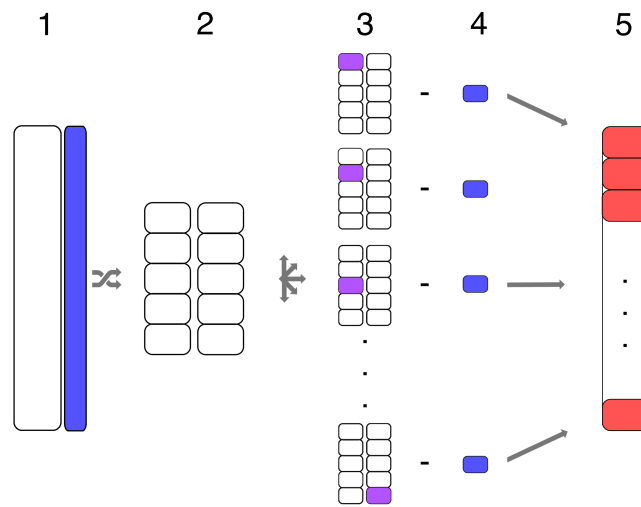


Figura 5.2: Cross-validation.

diretta per ottenere il punto corrispondente nello spazio e successivamente foveandolo con il sistema di visione (come è illustrato nella Sez. *Generazione dataset*). Per evitare di addestrare la rete neurale su una porzione limitata dello spazio, si sono riordinati casualmente i valori del punto 1 (mantenendo la corrispondenza fra di essi) e li si è divisi in 10 sottogruppi che chiameremo fold.

3. A partire dai fold generati si sono creati 10 set di dati, ognuno contenente 9 fold che verranno utilizzati per il training della rete neurale, ed un fold che non sarà incluso nei dati di training e sarà utilizzato per la simulazione (viola). Ripetiamo l'addestramento di ogni gruppo con le varie combinazioni di parametri e simuliamo la rete neurale ottenendo gli angoli del braccio.
4. Una volta calcolata la cinematica diretta degli angoli simulati, possiamo confrontarli con i punti 3D (blu) calcolati a partire dai dati in ingresso e derivare l'errore sotto forma di distanza euclidea. Stabiliamo così quali sono i parametri ottimali che minimizzano globalmente l'errore per ogni combinazione di fold scelta.

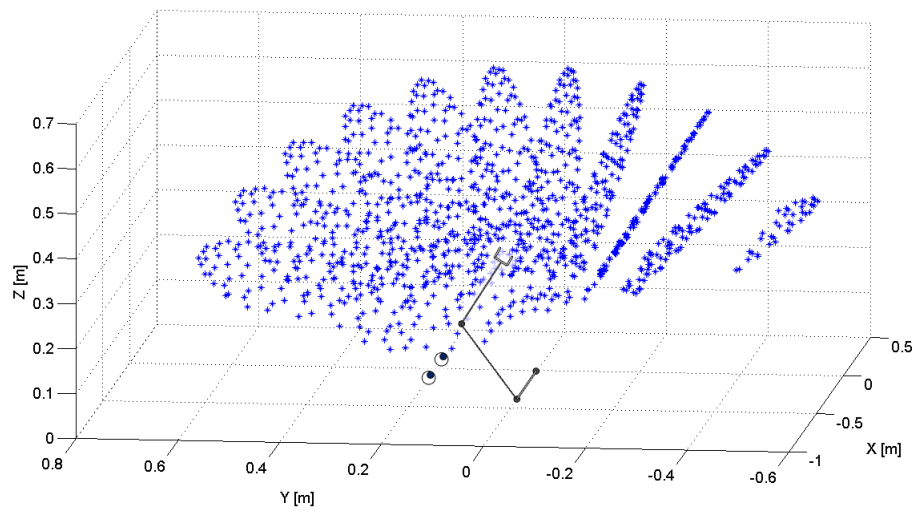


Figura 5.3: Spazio di lavoro del braccio robotico RRR. I vincoli introdotti negli angoli di giunto fanno sì che non si ottenga una semisfera piena come spazio di lavoro.

5. L'errore euclideo, corrispondente al processo di foveazione di ogni punto iniziale, è stato ottenuto eseguendo il calcolo della distanza fra il punto costruito dall'output simulato dalla rete neurale ed il punto 3D calcolato a partire dai dati in input alla rete. Per evitare che alcuni punti della simulazione fossero inclusi nel set di training e quindi provocassero una sottostima dell'errore, abbiamo sfruttato la suddivisione in dataset dei punti 3-4 e poi ricostruito ordinatamente il dataset d'errore.

Per applicare lo pseudo-algoritmo appena illustrato è stato usato un dataset di 1180 elementi, quindi ogni fold ha una dimensione 118 elementi. Complessivamente abbiamo addestrato la rete ad ogni ciclo con 1062 elementi e l'abbiamo validata con 118.

5.2.2 Braccio RRR

Introducendo un nuovo grado di libertà si è reso possibile il raggiungimento di un punto nello spazio 3D, tuttavia ora ogni punto è raggiungibile

da 4 configurazioni diverse del braccio. Abbiamo gestito la ridondanza imponendo al braccio robotico dei vincoli che ne mantenessero la postura il più vicino possibile ad una postura umana e che potesse essere percepita dall'uomo come umanoide, il giunto θ_0 della spalla ha un'escursione di $[-\frac{\pi}{2}, \frac{\pi}{2}]$, il giunto θ_1 della spalla di $[-\frac{\pi}{2}, \frac{\pi}{2}]$ ed il giunto θ_2 del gomito di $[0, \frac{3}{4}\pi]$. Lo spazio di lavoro raggiunto è mostrato in Fig. 5.3. Imponendo questi vincoli si elimina la ridondanza e si può passare alla generazione del dataset e al training della rete neurale, mantenendo lo stesso approccio illustrato nel braccio RR.

Capitolo 6

Risultati sperimentali

In questo capitolo si illustrano i risultati ottenuti con le tecniche precedentemente illustrate. Per prima cosa verranno presentati i risultati del sistema visivo, proponendo nuovi test per verificare e confrontare le caratteristiche del sistema base ed esteso. Successivamente si esporranno i risultati raggiunti con le reti neurali per il reaching, infine verrà considerata la sinergia del sistema di visione con il sistema di reaching.

6.1 Visione

6.1.1 Sistema base

La minimizzazione con discesa del gradiente della funzione di costo applicata al set di dati per l'addestramento, ha portato ai seguenti parametri:

$$K1 = 0.7076$$

$$K2 = 0.5257$$

$$K3 = 0.7136$$

È interessante osservare come la funzione di costo abbia molti minimi locali ma, dalle nostre osservazioni, essi non deteriorano le performance globali

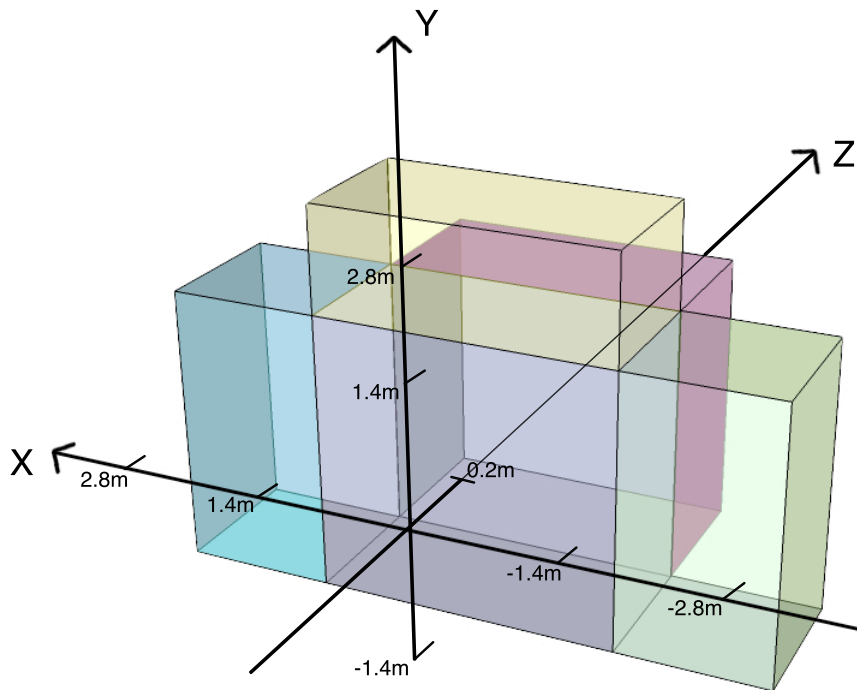


Figura 6.1: Sottospazi di testing. Il parallelepipedo blu centrale delimita lo spazio dei punti di addestramento. I parallelepipedi azzurro, giallo e rosso delimitano lo spazio su cui viene eseguito il test di esplorazione dello spazio 3D. In verde lo spazio su cui si esegue il test di posizione iniziale.

del sistema. Per testare le prestazioni del sistema di controllo, si conducono gli esperimenti di esplorazione dello spazio 3D e test della posizione iniziale.

Esplorazione dello spazio 3D

Indaga la capacità del sistema di stereo-visione di foveare punti non contenuti nel set di addestramento partendo ogni volta dalla stessa posizione iniziale (posizione di riposo con $\theta_r = \theta_l = \theta_t = 0$). È possibile raggruppare i punti che si testeranno in tre sottospazi adiacenti allo spazio dei punti dell'addestramento, basandosi sulla loro posizione nello spazio 3D

(interpretazione grafica in Fig. 6.1):

Lungo la direzione z : $[-1.4, 1.4] \times [-1.4, 1.4] \times [1.4, 2.8]$ metri

Lungo la direzione x : $[1.4, 2.8] \times [-1.4, 1.4] \times [0.2, 1.4]$ metri

Lungo la direzione y : $[-1.4, 1.4] \times [1.4, 2.8] \times [0.2, 1.4]$ metri

Ognuna di queste porzioni di spazio è discretizzata con un passo di 0.01 metri in ogni direzione. Vengono esclusi i punti che si proiettano in uno solo dei due piani immagine, si studierà questa situazione nella Sez. *Alieno*.

L'errore medio associato ai punti di test lungo la direzione Z è di 1.60 pixel con una varianza di 0.60. Ci si aspettava questo risultato principalmente perché le proiezioni dei punti 3D sono più vicine al centro dell'immagine al crescere della loro distanza dal piano immagine. Lungo la direzione delle X l'errore cresce al crescere di x. Siccome questi punti sono vicini al piano immagine le loro proiezioni sono nei bordi delle immagini e, conseguentemente, la foveazione di essi è più sfidante. Ad ogni modo, come si può osservare nella parte inferiore della figura, l'errore è distribuito in un intervallo accettabile con una media di 3.35 pixel e una varianza di 0.53. Si possono fare considerazioni simili per i punti di testing lungo la direzione Y, dove qualitativamente l'errore cresce col crescere della componente y del punto 3D. La media è di 3.33 pixel e la varianza di 0.76 (Fig. 6.2).

Test della posizione iniziale

Indaga la capacità del sistema di foveare una feature nello spazio 3D, indipendentemente dalla posizione iniziale delle telecamere. L'obiettivo è verificare la robustezza del sistema nel foveare un punto a partire da una generica posizione iniziale degli angoli θ del sistema di visione. Vergenza e versione influiscono antagonisticamente sul comando di panning. Per verificare se il sistema sia in grado di effettuare il panning correttamente valutiamo la regione più problematica dello spazio 3D. La regione Z rappresenta un caso "facile", infatti i punti sono proiettati sempre al centro

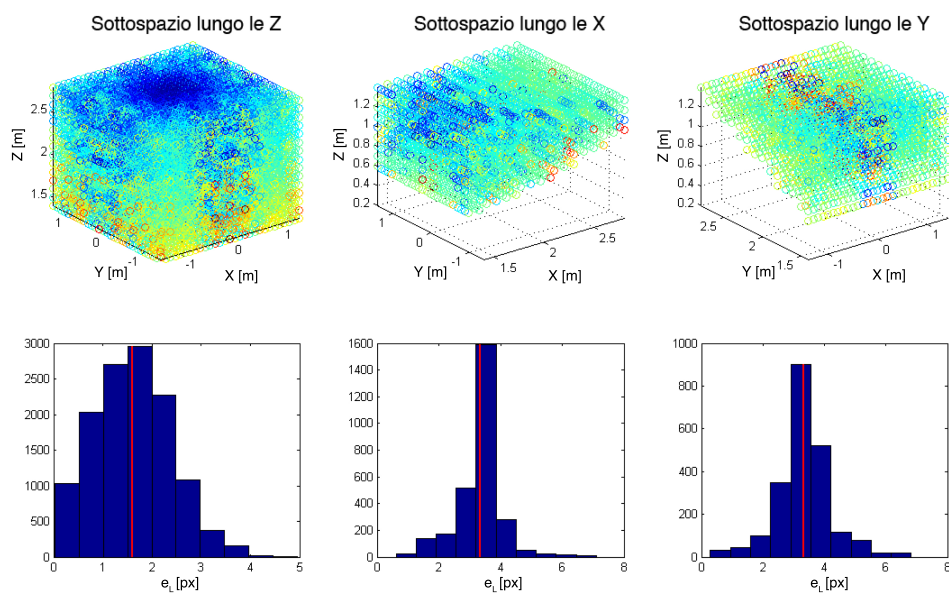


Figura 6.2: Risultati test esplorazione dello spazio 3D per il sistema base. Mappa dell'errore calcolata per l'occhio sinistro; abbiamo osservato errori molto simili anche per l'occhio destro. Riga superiore: set di test con errore associato ad ogni punto 3D foveato. Riga inferiore: distribuzione dell'errore in pixel per ogni set di test. La linea rossa rappresenta la media dell'errore.

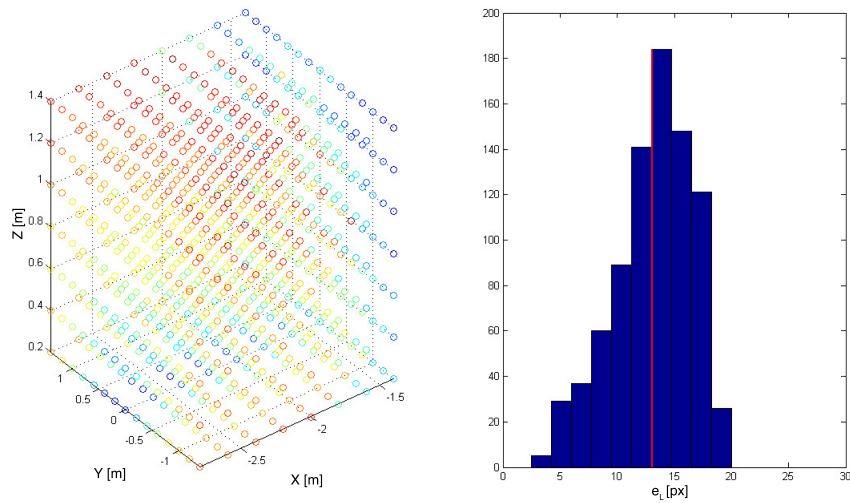


Figura 6.3: Risultati test posizione iniziale per il sistema base. Nella parte sinistra dell'immagine è mostrato l'errore medio associato ad ogni punto 3D del set di test. L'errore medio è calcolato considerando ogni posizione iniziale plausibile dei giunti della testa; per ogni punto calcoliamo l'errore per la foveazione. Nell'immagine di destra è mostrata la distribuzione dell'errore medio..

dell'immagine e la regione Y non agisce sul panning bensì sul tilting. Il sottospazio utilizzato per l'esperimento è quindi lungo la direzione X:

$$[-2.8, -1.4] \times [-1.4, 1.4] \times [0.2, 1.4] \text{ metri}$$

discretizzato con un passo di 0.2 metri su ogni direzione. Si procede facendo foveare al sistema ognuno dei punti di test a partire da ogni possibile configurazione dei giunti. Definiamo un range di valori per ogni giunto, i.e. [- 60 gradi, 60 gradi] con uno step di 30 gradi. In totale si hanno 125 diverse configurazioni dei giunti, di cui ne vengono considerate solo 75 scartando quelle che portano il sistema ad una situazione iniziale non accettabile (i.e. telecamere divergenti che corrisponde ad avere occhi strabici). Infine si calcola la media dell'errore associata ad ogni punto 3D ed i risultati sono mostrati in Fig. 6.3. Mediamente si ha un errore di 13.10 pixel distribuito nell'intervallo di valori [2.47, 20.01] pixel. Qualitativamente

l'errore aumenta all'allontanarsi della x e della z dall'origine del sistema, dovuto principalmente alla posizione marginale del sottospazio preso in considerazione.

6.1.2 Sistema esteso

Gli esperimenti presentati in questa sezione si propongono di:

- selezionare la configurazione del collo che offre le migliori performance in termini di errore nell'esplorazione dello spazio 3D di testing;
- mettere a confronto gli esiti di questi test, con quelli del sistema base senza collo mobile.

Per selezionare la migliore configurazione abbiamo confrontato vicendevolmente i risultati ottenuti nell'esperimento di esplorazione dello spazio 3D. La configurazione migliore è stata poi utilizzata per eseguire il test sulla posizione iniziale.

Esplorazione dello spazio 3D

Eseguiamo l'esperimento come descritto nel caso analogo del sistema di base (Sezione 6.1.1) per ogni configurazione del sistema esteso riportata nella Tabella 4.1. Confrontando i risultati di media e deviazione standard abbiamo stabilito che la configurazione che permette di minimizzare l'errore è la quinta. È interessante osservare che un controllore di questo tipo ha anche una certa plausibilità biologica, infatti Chen ha dimostrato tramite una serie di esperimenti di elettrostimolazione su scimmie, che il movimento di collo e occhi può essere controllato in maniera indipendente sia volontariamente che involontariamente [35]. I range di valori sul quale è stato eseguito l'esperimento sono gli stessi utilizzati per il test analogo sul

sistema base:

$x \in [-1.4, 1.4]$ metri con step di 0.7 metri

$y \in [-1.4, 1.4]$ metri con step di 0.7 metri

$z \in [0.2, 1.4]$ metri con step di 0.4 metri

I parametri K ottenuti dopo la fase di addestramento della configurazione 5 sono:

$$K1 = 0.0199$$

$$K2 = 0.3770$$

$$K3 = 0.7079$$

$$K4 = 0.6706$$

che conducono ai risultati illustrati in Fig. 6.4. Tali risultati sono simili alle performance ottenute nel sistema base senza collo mobile (vedi Sezione 6.1.1 e Fig. 6.2); nello specifico otteniamo una media di 3.38, 3.03, 1.46 pixel con relativa varianza di 0.12, 0.12, 0.30 rispettivamente per i test effettuati nel sottospazio lungo X,Y e Z.

Test posizione iniziale

Analogamente alla Sezione 6.1.1 abbiamo eseguito il test della posizione iniziale, in questo caso su 375 configurazioni iniziali, ottenendo un errore distribuito nell'intervallo di valori [3.14, 9.31] pixel con media 5.23 pixel, Fig. 6.5.

6.1.3 Confronto fra sistema base e sistema esteso

Come è possibile notare dai risultati ottenuti nei test (riportati per comodità nella Tabella 6.1), l'introduzione di un ulteriore grado di libertà nel

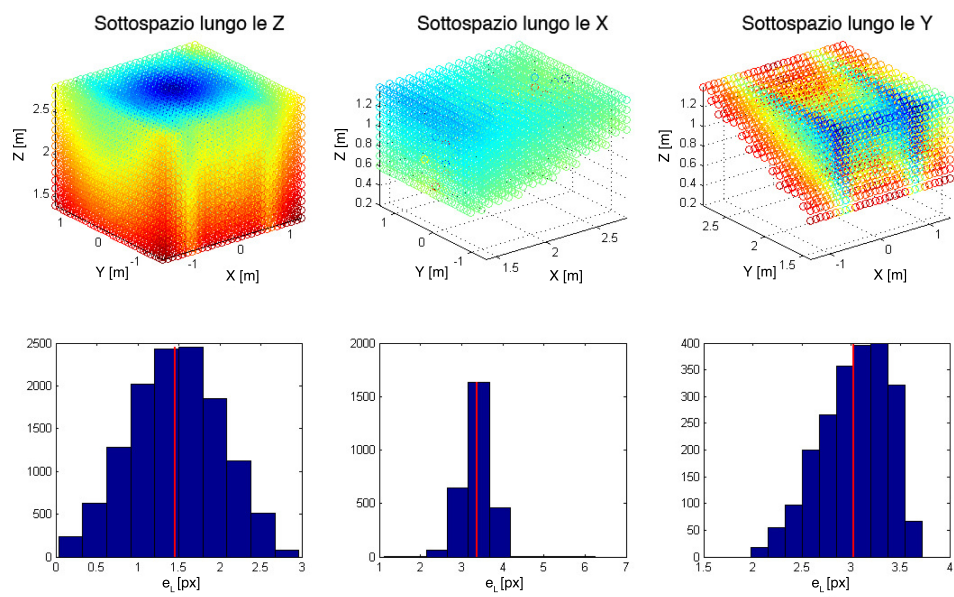


Figura 6.4: Risultati test esplorazione dello spazio 3D per il sistema esteso. Mappa dell'errore calcolata per l'occhio sinistro; abbiamo osservato errori molto simili anche per l'occhio destro. Riga superiore: set di test con errore associato ad ogni punto 3D foveato. Riga inferiore: distribuzione dell'errore in pixel per ogni set di test. La linea rossa rappresenta la media dell'errore.

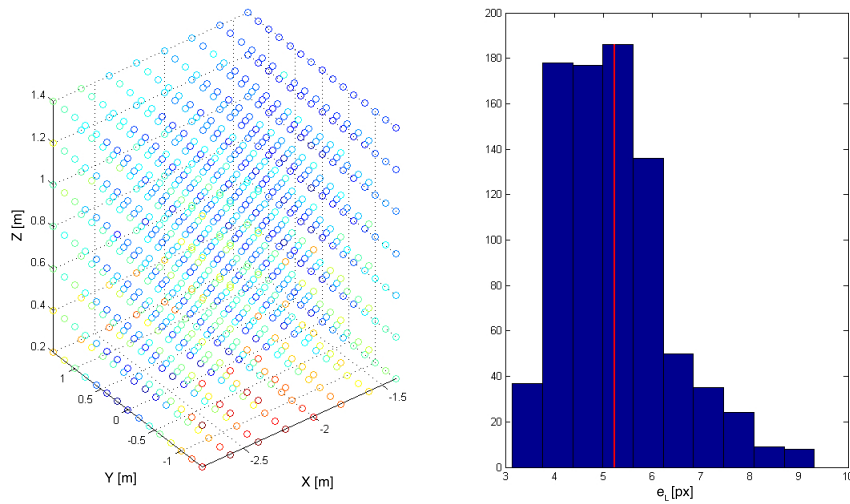


Figura 6.5: Risultati test posizione iniziale per il sistema esteso. Nella parte sinistra dell'immagine è mostrato l'errore medio associato ad ogni punto 3D del set di test. L'errore medio è calcolato considerando ogni posizione iniziale plausibile dei giunti della testa; per ogni punto calcoliamo l'errore per la foveazione. Nell'immagine di destra è mostrata la distribuzione dell'errore medio.

sistema non comporta un peggioramento delle prestazioni, anzi, mediamente a fronte di una modesta variazione della media dell'errore rispetto al caso base, si osserva una riduzione della varianza. Questo porta ad ottenere dei risultati più stabili e grafici del mapping dell'errore più "puliti". È interessante osservare come nel grafico del sottospazio lungo l'asse Y, i punti con minor errore si siano spostati: la loro concentrazione non è più maggiore intorno allo zero dell'asse X, ma si concentra in due fasce. Questa variazione di distribuzione è dovuta al fatto che l'aggiunta del grado di ridondanza nel controllore porta ad avere dei fenomeni di vibrazione del sistema nel caso i punti da foveare siano esattamente di fronte al robot: il controllore infatti dovrebbe effettuare solo il panning degli occhi per la convergenza senza muovere il giunto del collo come avviene correttamente nel sistema base.

Per quanto riguarda il test della posizione iniziale, confrontato con le

performance del sistema base a tre gradi di libertà, l'errore presenta una media e deviazione standard notevolmente minori. Possiamo quindi concludere che il giunto del collo aggiunto fornisce una maggiore robustezza al sistema e, nello specifico, riduce l'influenza della configurazione iniziale della testa nelle prestazioni di foveazione di un punto nello spazio.

<i>Test</i>	<i>Sistema base</i>	<i>Sistema esteso</i>
Esplorazione sottospazio X	$\mu = 3.35 \text{ pixel}$ $\sigma^2 = 0.53$	$\mu = 3.38 \text{ pixel}$ $\sigma^2 = 0.12$
Esplorazione sottospazio Y	$\mu = 3.33 \text{ pixel}$ $\sigma^2 = 0.76$	$\mu = 3.03 \text{ pixel}$ $\sigma^2 = 0.12$
Esplorazione sottospazio Z	$\mu = 1.60 \text{ pixel}$ $\sigma^2 = 0.60$	$\mu = 1.46 \text{ pixel}$ $\sigma^2 = 0.30$
Test posizione iniziale	$\mu = 13.10 \text{ pixel}$ $\sigma^2 = 11.72$	$\mu = 5.23 \text{ pixel}$ $\sigma^2 = 1.22$

Tabella 6.1: Riassunto dei risultati ottenuti nei test sul sistema base ed esteso.

Errore nel 2D e nel 3D

Si è scelto di utilizzare per l'apprendimento dei parametri K e per la verifica dell'errore, la distanza euclidea calcolata nel piano immagine fra la posizione finale della feature proiettata ed il centro del piano immagine medesimo (in questo caso abbiamo l'origine del sistema di riferimento del piano immagine esattamente al centro del piano immagine stesso). La distanza euclidea usata nella funzione obiettivo è necessaria a valutare la performance del sistema nel foveare il punto desiderato. Questa interpretazione dell'errore è perfettamente compatibile con un approccio di matrice umana, infatti il movimento di foveazione dell'occhio umano è finalizzato

proprio a minimizzare la distanza fra proiezione dell'obiettivo sulla retina e centro della retina, che è il punto con maggiore densità di recettori.

Si ritiene comunque interessante stimare l'errore nello spazio 3D in modo da averne un'idea sia qualitativa che quantitativa. Si calcola quindi la distanza euclidea fra la feature originale e il punto foveato dal sistema, tale punto è ricostruito con la triangolazione lineare delle sue proiezioni nelle due viste e le relative matrici di proiezione. In ogni piano immagine infatti, abbiamo le coordinate $x = PX, x' = P'X$ con X punto 3D. Queste equazioni possono essere combinate nella forma $AX = 0$, che è un sistema di equazioni lineari in X risolvibile tramite i metodi di calcolo numerico (per una maggiore comprensione si rimanda alla Sezione 2.3.1).

Nel sistema base abbiamo ottenuto un errore 3D medio di 6.61 centimetri con scarto quadratico medio di 2.9 centimetri, mentre nel sistema esteso otteniamo un leggero miglioramento avendo un errore 3D medio di 6.08 centimetri con uno scarto di 2.7 centimetri. Nella Fig. 6.6 abbiamo riportato il piano XY dove ad ogni punto visibile è associato l'errore 3D commesso nella foveazione, sopra per il sistema base e sotto per il sistema esteso. Anche in questo caso è possibile apprezzare come l'introduzione del collo porti ad una distribuzione dell'errore più pulita e lineare. Nella Fig. 6.7 sono stati disegnati in blu la feature da foveare e in azzurro quella effettivamente foveata dalle telecamere: in entrambi i casi si evince che l'errore 3D è dovuto principalmente ad un'errata vergenza che porta a foveare un punto con una distanza, nella maggior parte dei casi, inferiore a quella desiderata. Mentre nel sistema base tale errore è ancora legato alla posizione x del punto (si noti come la stima del punto peggiori in modo indeterminabile e casuale per $|x| > 0.5$), nel sistema esteso è possibile riconoscere un pattern legato all'errore: la distribuzione dei punti azzurri è infatti regolare e radiale rispetto al centro del sistema, quindi meno dipendente dalla x ma proporzionale alla distanza euclidea fra feature e origine del mondo.

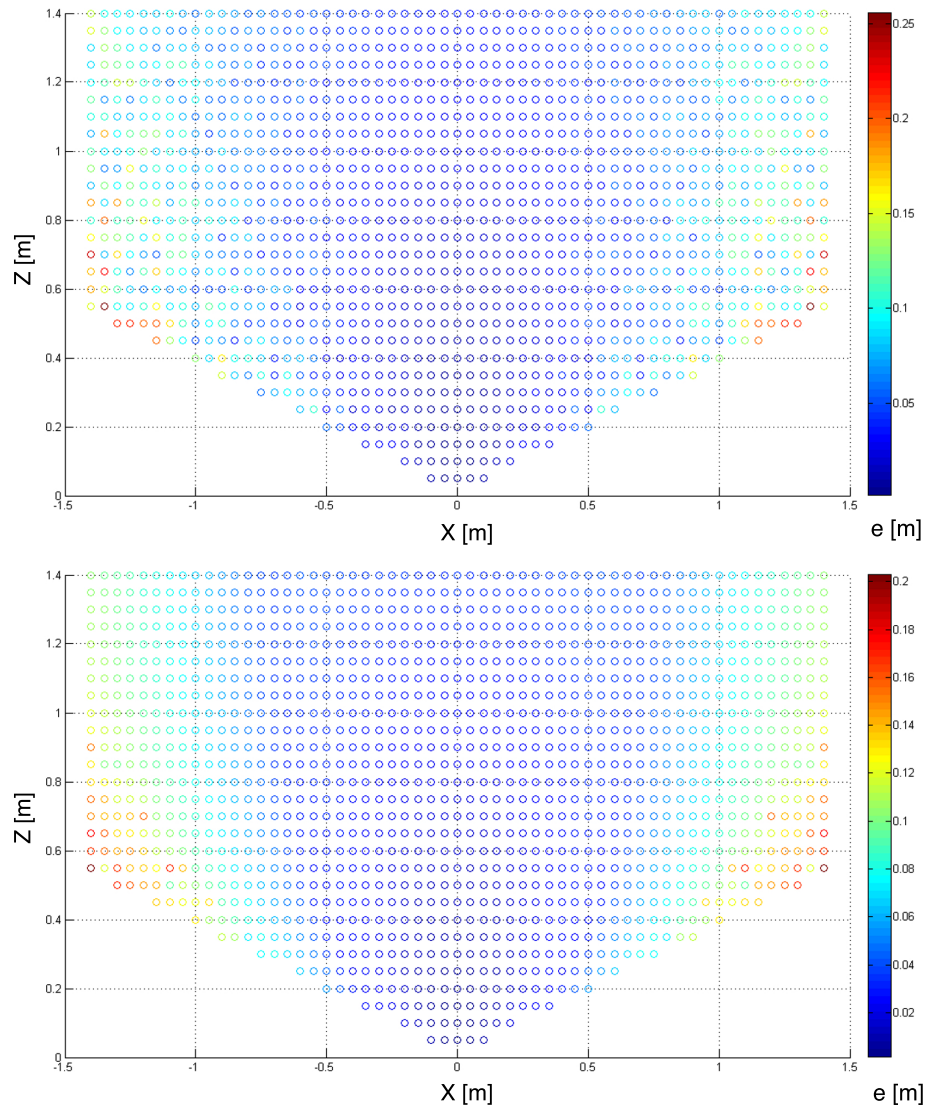


Figura 6.6: Nella parte superiore dell'immagine è mostrato l'errore nel sistema base e in quella inferiore l'errore nel sistema esteso. L'errore 3D è associato al punto da foveare nel piano XZ.

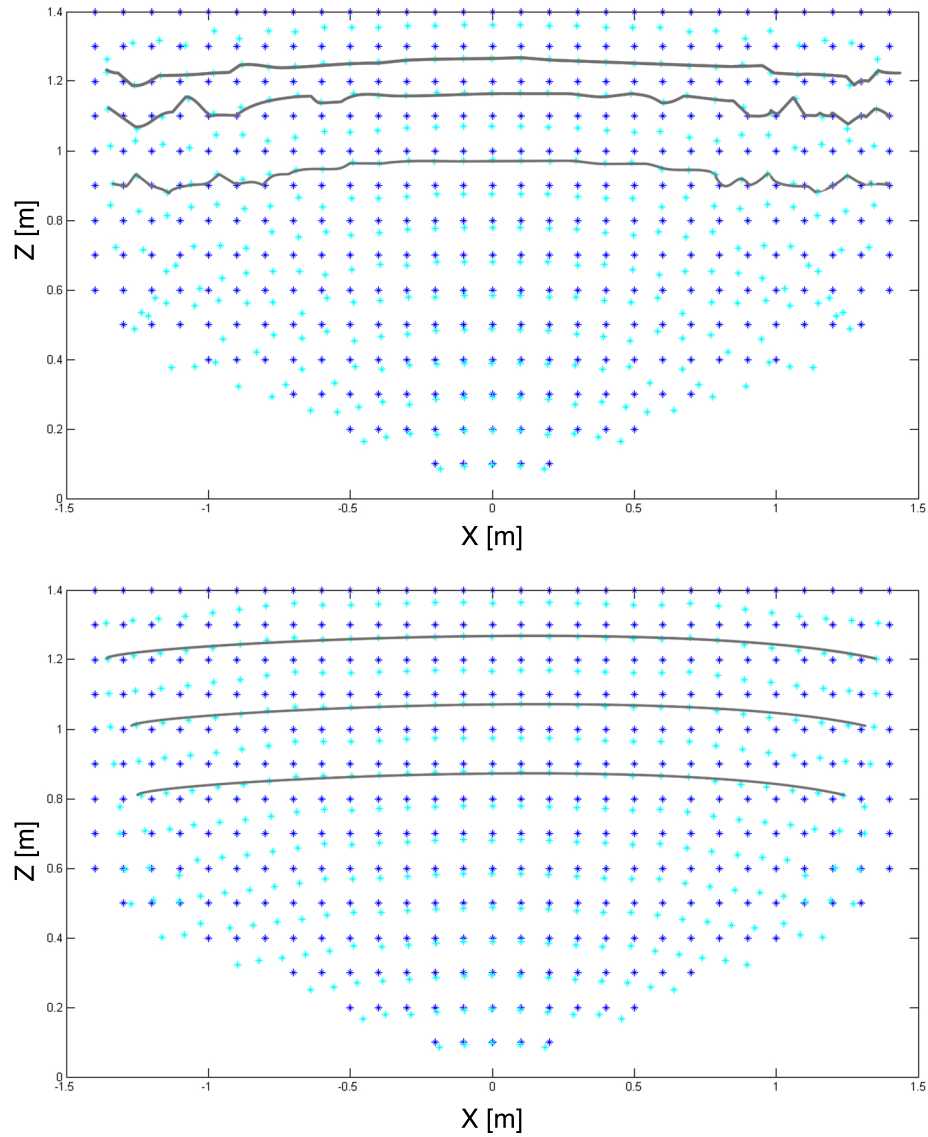


Figura 6.7: Nella parte superiore dell'immagine è mostrato l'errore nel sistema base e in quella inferiore l'errore nel sistema esteso. Vengono rappresentati in blu i punti da foveare e in azzurro i punti effettivamente foveati dai rispettivi sistemi.

In Fig. 6.8 abbiamo scelto di rappresentare visivamente la distanza che separa la feature dal punto effettivamente foveato per ogni sistema (base sulla sinistra, esteso sulla destra) e per ogni piano (XZ, YZ e XY procedendo dall'alto). Come si può osservare i segmenti degli errori, se prolungati, tendono al centro, precisamente verso il punto di coordinate $[0, 0.2, 0]$ che corrisponde all'occhio ciclopico virtuale. L'errore di foveazione è quindi solo sull'asse che congiunge la feature all'occhio ciclopico. L'effetto "ventaglio rosso" è dovuto alla proiezione di tutti i segmenti dello spazio 3D su un piano. Confrontando la parte sinistra con la destra si ha un'ulteriore prova della riduzione della variabilità dell'errore con l'introduzione del collo, infatti i ventagli di destra sono meno ampi, quasi un'unica linea. Da evidenziare soprattutto è il fenomeno che si osserva nei ventagli del piano XY del sistema esteso che si riducono ad un solo segmento, ciò è dovuto ad una notevole riduzione della varianza dell'errore lungo gli assi X e Y, grazie all'introduzione del collo la componente principale dell'errore è lungo l'asse Z.

Nella Fig. 6.9 abbiamo riportato in verde i punti in cui l'errore 3D di foveazione del sistema esteso è minore del sistema base, in rosso quando invece il punto viene stimato meglio dal sistema base. Evidente è la fascia centrale rossa, ulteriore prova del peggioramento delle performance del sistema esteso dovuto a fenomeni di vibrazione per i punti che si trovano esattamente di fronte al sistema.

Per discriminare ulteriormente la diversità di comportamento che abbiamo appena descritto, abbiamo calcolato la differenza fra gli errori di foveazione dei punti del sistema base ed esteso e l'abbiamo rappresentata in Fig. 6.10 dove abbiamo contrassegnato in rosso i punti in cui il sistema con rotazione del collo ha un comportamento peggiore di quello del sistema senza rotazione. Nella barra laterale è riportato il gradiente di colore che viene associato al range di valori che assume la differenza di errori; si

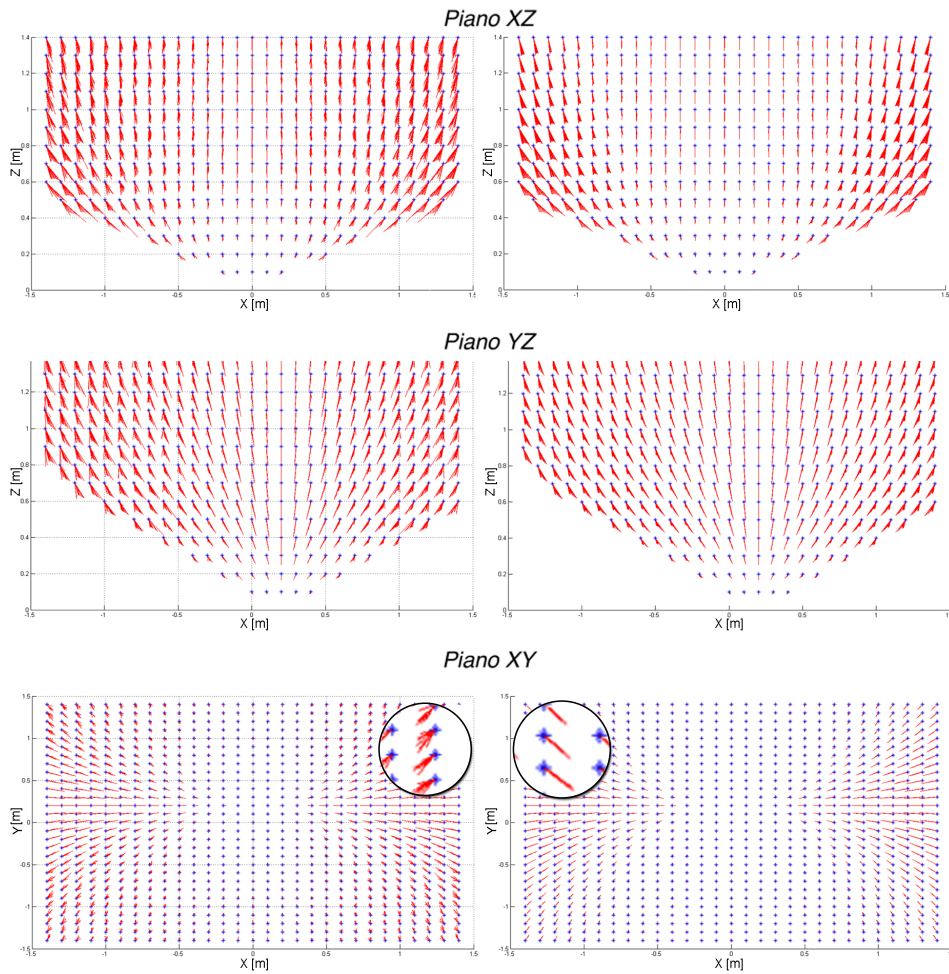


Figura 6.8: Nella parte sinistra dell'immagine è mostrato l'errore nel sistema base e in quella destra l'errore nel sistema esteso. In blu sono rappresentate le feature da foveare, la distanza fra esse e i relativi punti foveati è rappresentata in rosso.

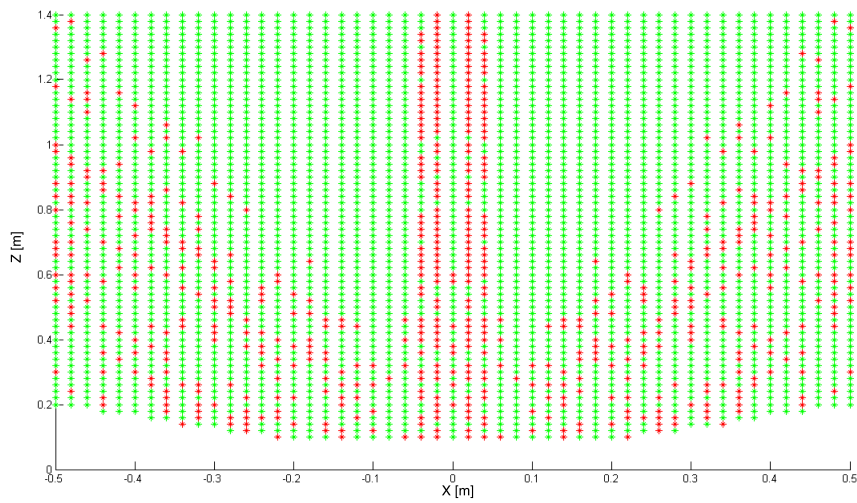


Figura 6.9: In verde i punti in cui l'errore di foveazione del sistema esteso è minore rispetto a quello del sistema base.

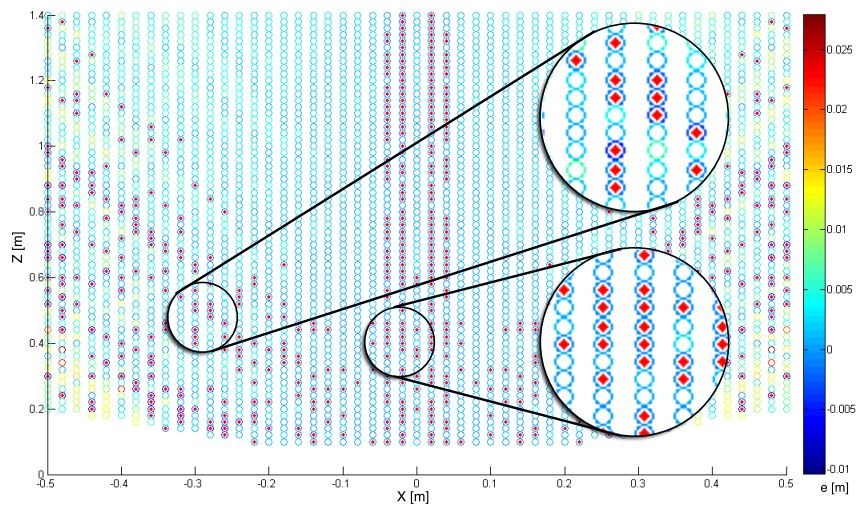


Figura 6.10: Confronto delle performance dei due sistemi di visione.

noti come lo zero non è centrale ma sia spostato in basso, segno che il magnitudo dei pochi casi in cui il sistema base ha delle performance migliori è notevolmente inferiore dei casi in cui il sistema esteso fovea meglio. Di conseguenza si può affermare che in linea generale il sistema esteso ha prestazioni migliori e al più simili a quelle ottenute dal sistema base.

6.1.4 Osservazioni

In questa sezione esporremo alcune considerazioni riguardanti il sistema visivo esteso le quali, essendo di secondaria importanza, non hanno trovato spazio in precedenza. Tuttavia vale la pena riportarle in quanto hanno permesso di raggiungere conclusioni degne di nota.

Collo autovincolato

Il nostro sistema di visione composto dalle due telecamere e dal collo che può eseguire movimenti di pan e tilt, permette la foveazione di un punto con diverse configurazioni di angoli di giunto avendo più gradi di libertà di quelli strettamente necessari per il compito (sistema ridondante). Come è possibile quindi che il nostro sistema riesca a scegliere sempre la configurazione human-like? La risposta è intrinseca alla funzione del controllore da minimizzare: ponendo come obiettivo la riduzione della distanza della feature proiettata dal centro dell'immagine retinica, il controllore automaticamente vincola anche i movimenti di occhi e collo, non permettendo rotazioni che porterebbero il sistema in condizioni umanamente impossibili.

Traiettorie

È interessante effettuare un'analisi qualitativa delle traiettorie di foveazione tracciate dagli angoli degli occhi e del collo per fissare un oggetto (Fig. 6.11). Per evidenziare il ruolo del collo nel panning rispetto al movimento degli occhi, abbiamo considerato solo casi in cui non era coinvolto il tilting (punti ad altezza delle telecamere). Inizialmente collo e telecamere si trovano in posizione di riposo. Nel momento in cui le telecamere si trovano a dover foveare un punto (contrassegnato con una croce blu nell'immagine), si assiste ad un rapido (0.03 secondi) movimento di panning del collo composto a un piccolo movimento di versione delle telecamere verso la posizione desiderata, seguito poi da un movimento più lento e preciso

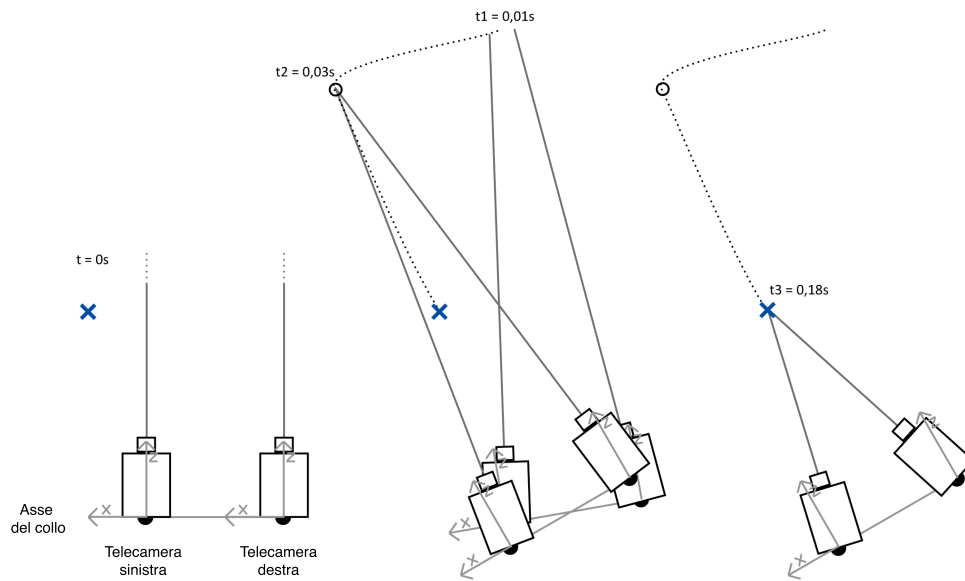


Figura 6.11: Sono illustrate le traiettorie compiute dalle telecamere nel sistema esteso. La croce blu rappresenta la future 3D di coordinate [200, 0, 40] che deve essere raggiunta nello spazio. Le immagini sono state scalate per mostrare più chiaramente che il sistema prima muove il collo, solo quando il collo ha quasi raggiunto la sua posizione definitiva viene eseguito il movimento di vergenza.

di convergenza delle telecamere (0.18 secondi) verso il punto. Questo tipo di comportamento è sovrapponibile a quello umano, anch'esso diviso in una prima fase di panning più veloce e grezza eseguita principalmente dal collo, seguita da una più precisa di convergenza.

Alieno

Illustriamo un ulteriore esperimento che abbiamo eseguito. Esso si discosta leggermente dall'approccio ispirato all'essere umano, ma abbiamo ritenuto utile eseguirlo per indagare meglio il funzionamento del sistema di controllo. Implementando il sistema human inspired di telecamere e collo illustrato in precedenza, si era notato che a causa della dimensione del campo visivo di un singolo occhio e della breve distanza fra i due occhi, i punti visti da un singolo occhio erano esigui e quindi un campione troppo ridotto per effettuare dei test.

Il nostro obiettivo è verificare se l'introduzione del collo, oltre a rendere più uniforme l'errore nel campo visivo di entrambi gli occhi, rendesse foveabili con un errore plausibile anche i punti che prima del movimento del collo stesso erano visibili a un singolo occhio. Abbiamo quindi modificato il modello del robot ampliando il campo visivo di entrambi gli occhi, distanziandoli maggiormente in modo da avere uno spazio periferico del campo visivo visibile da un solo occhio di dimensioni maggiori. Il campo visivo di un singolo occhio passa quindi ora a 145 gradi e la distanza interpupillare a 40 centimetri. Come per il modello del sistema esteso, abbiamo minimizzato nuovamente la funzione di costo, calcolato i parametri caratteristici del controllore ed eseguito il testing. I parametri K risultanti sono:

$$K1 = 0.0500$$

$$K2 = 0.1617$$

$$K3 = 0.6534$$

$$K4 = 0.3760$$

Dai test sovrastanti e dalla rappresentazione dell'errore 3D di foveazione (Fig. 6.12), si può notare che la distribuzione dell'errore nell'area visibile da un solo occhio tende ad aumentare, è però strettamente correlata all'errore del campo di visione comune ad entrambi gli occhi e si raccorda con

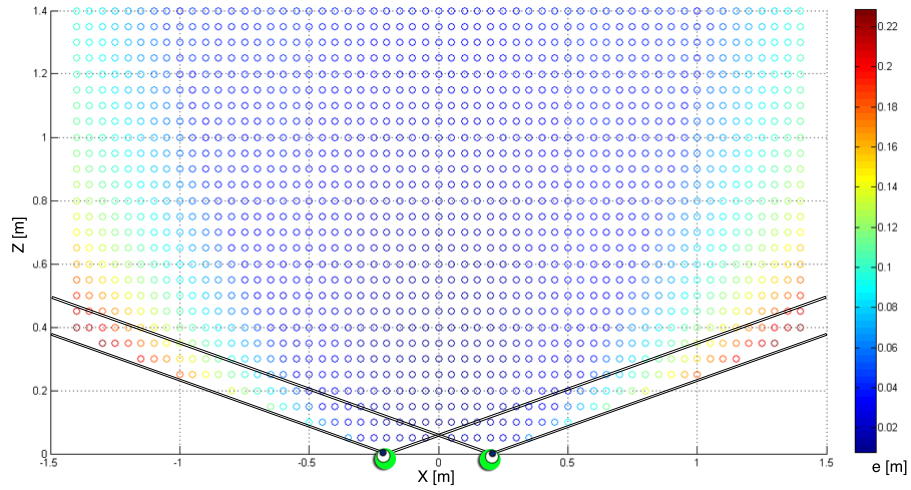


Figura 6.12: Campo visivo dell'alieno. Abbiamo aggiunto delle linee nere che indicano l'apertura del campo visivo di ogni singolo occhio per evidenziare i punti che cadono nello spazio periferico.

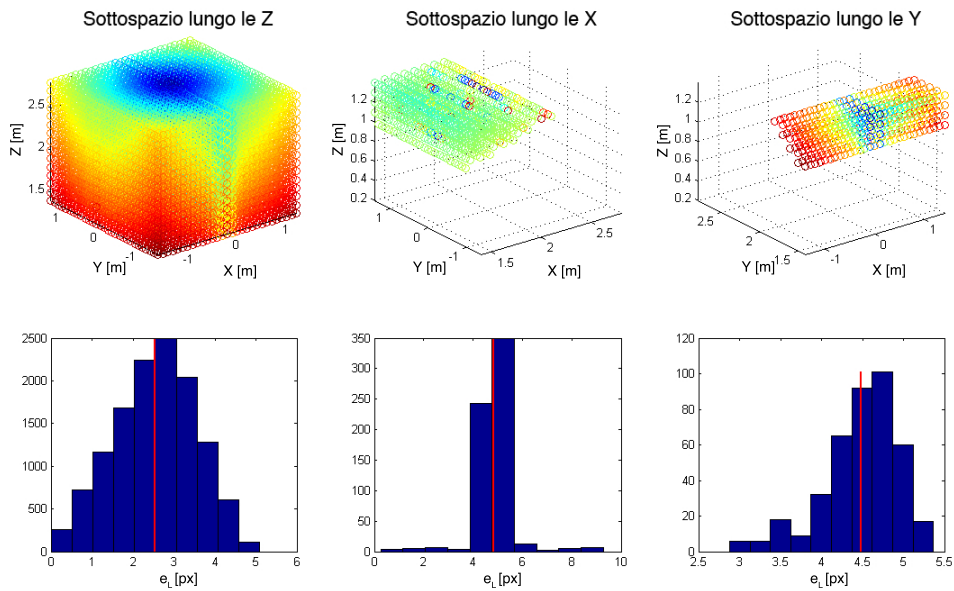


Figura 6.13: Mappa dell'errore nei test di esplorazione dello spazio 3D relativi al caso "alieno".

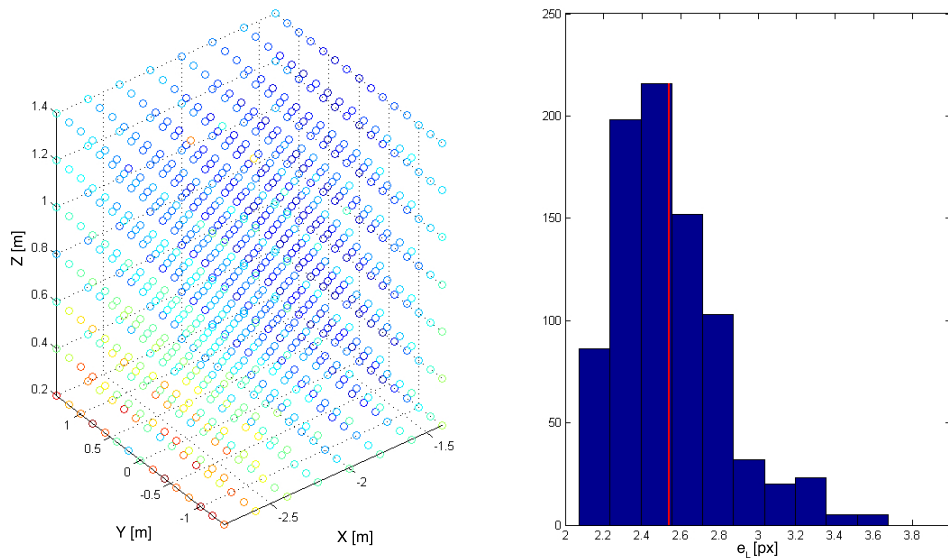


Figura 6.14: Test posizione iniziale del caso "alieno".

esso in modo continuativo. Ciò implica che l'errore di questi punti è legato alla posizione nello spazio e non al fatto che sono inizialmente visibili solo da un occhio. Considerando inoltre la posizione di quei punti ed il fatto che questo è un esperimento che rappresenta un ramo morto all'interno dello sviluppo del progetto e dal quale non verranno calcolati ulteriori dati, un errore nel 3D di 5.68 centimetri con scarto quadratico medio di 4 centimetri, può essere ritenuto soddisfacente. Questo è quindi un ulteriore fattore che giustifica l'introduzione della rotazione nel giunto del collo (si vedano anche i risultati dei test esplorazione 3D e test posizione iniziale mostrati in Fig. 6.13 e Fig. 6.14). È interessante osservare che, nell'area del campo visivo comune ai due occhi, l'errore si riduce rispetto al caso in cui il robot rispecchiava i parametri umani degli occhi; questo accade perché maggiore è la distanza fra le telecamere, definita baseline, migliore è la precisione e stima della distanza degli oggetti dal soggetto. La baseline è infatti il parametro più importante di progetto di un sistema stereoscopico.

6.2 Reaching

6.2.1 Braccio RR

Implementando lo pseudo-algoritmo illustrato in *Rete neurale* della Sezione 5.2.1 per ottenere i parametri caratteristici di una rete neurale, abbiamo determinato come valore ottimale dello spread $0.55[\text{rad}]$ e della tolleranza 0.001 , essi portano a posizionare mediamente 120 centri (Fig. 6.15). L'errore del compito di reaching compiuto dal braccio RR controllato dalla rete neurale ha una media di 0.0047 metri e scarto quadratico medio di 0.0151 metri, come mostrato in Fig. 6.16. Questo valore è ottenuto componendo gli errori di ognuno dei sotto-set di dati utilizzati per la Cross-validation, è quindi una stima molto affidabile della capacità della rete di generalizzare una soluzione a partire da valori in ingresso non inclusi nel set di training.

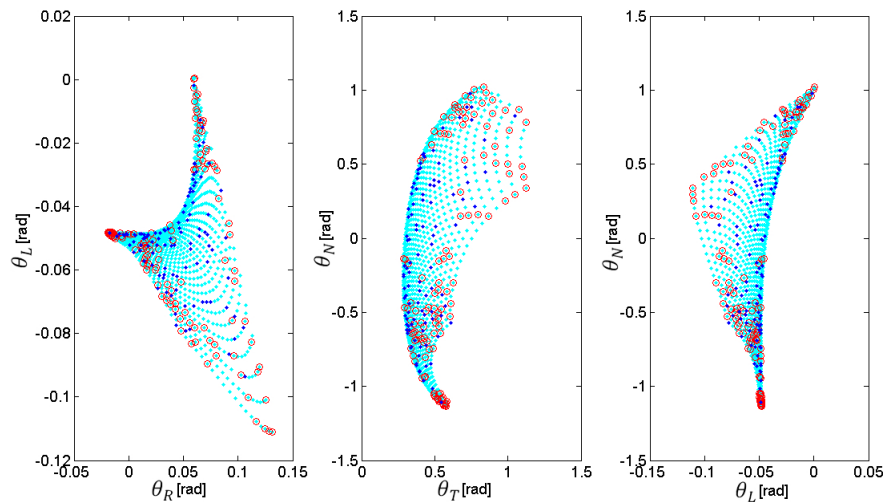


Figura 6.15: Posizionamento dei centri in corrispondenza dei valori in ingresso del sistema visivo. In azzurro i punti di training, in rosso i centri, in blu i valori casuali scelti per la simulazione.

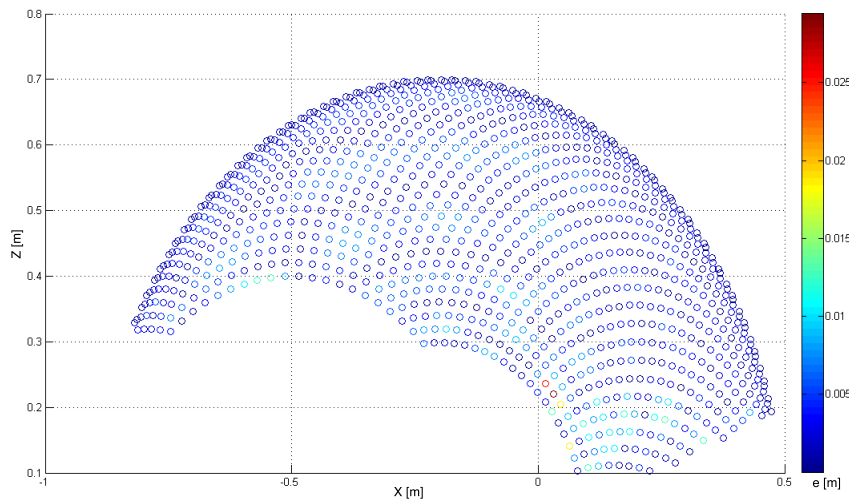


Figura 6.16: Distribuzione dell'errore di reaching nel braccio RR.

6.2.2 Braccio RRR

Abbiamo proseguito implementando lo pseudo-algoritmo anche nel caso del braccio RRR ed abbiamo determinato come valori ottimali dello spread 1.25[rad] e della tolleranza 0.0175, essi portano a posizionare 82 centri. L'errore del compito di reaching compiuto dal braccio RRR controllato dalla rete neurale, ha una media di 0.0320 metri e scarto quadratico medio di 0.0591 metri. Anche in questo caso si ha il riscontro della capacità della rete di approssimare una funzione e di generalizzare in maniera ottima anche per quei valori in ingresso non inclusi nel set di training.

Come si era illustrato nella sezione del capitolo di background riguardante le reti neurali (Sez. 2.4.2), le reti RBF possono approssimare le funzioni lineari con una precisione desiderata. Nel nostro caso però, i valori degli angoli del braccio che la rete neurale dovrà approssimare non sono rappresentabili con una funzione continua, infatti osservando la Fig. 6.17 si può notare come i valori assunti dagli angoli dei giunti del braccio RRR creino dei grafici costituiti da linee spezzate (in blu). θ_0 e θ_1 rappresentano i giunti di rotazione della spalla, θ_2 del gomito; passando dal "focus" 1 al 4, si può

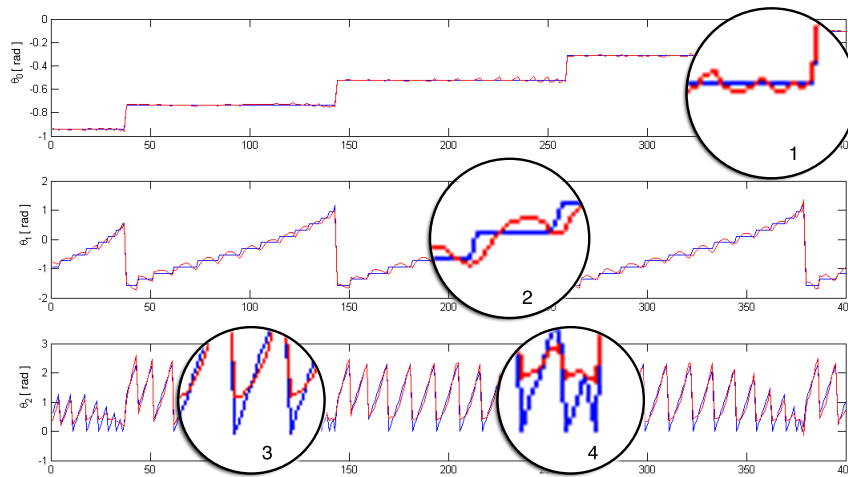


Figura 6.17: In blu i valori assunti dai giunti θ_0 , θ_1 e θ_2 , rispettivamente i due giunti della spalla e del gomito, su una porzione del dataset. In rosso i valori simulati dalla rete neurale.

osservare come la rete neurale (in rosso) introduca un errore maggiore nella stima dei valori di uscita all'aumentare del numero di variazioni repentine di direzione dei dati in ingresso. In 1 la stima è abbastanza precisa con qualche disturbo che non introduce un errore eccessivo in prossimità degli scalini (verificato sperimentalmente). In 2 si osserva in modo distinto che i gradini sono approssimati con delle gaussiane, a volte sovrastimando, altre sottostimando i valori in uscita. La situazione peggiora in 3 e 4: nei picchi verso il basso la rete non riesce a produrre in output i valori che dovrebbe simulare, approssimandone a volte anche in modo grezzo l'andamento. Il caso del braccio RR è analogo, raggiungendo però solo uno spazio lineare, è costituito da soli due giunti di rotazione. I loro grafici saranno qualitativamente simili a quelli del giunto θ_0 e θ_1 del braccio RRR, ma non verrà introdotto l'errore del terzo giunto, producendo quindi una simulazione più precisa.

Abbiamo notato che è presente un elemento con un errore molto alto: 0.9663 metri, esso è associato ad un punto che si trova molto vicino al bordo dello spazio di lavoro, nell'estremo di esso più vicino alla testa (si veda

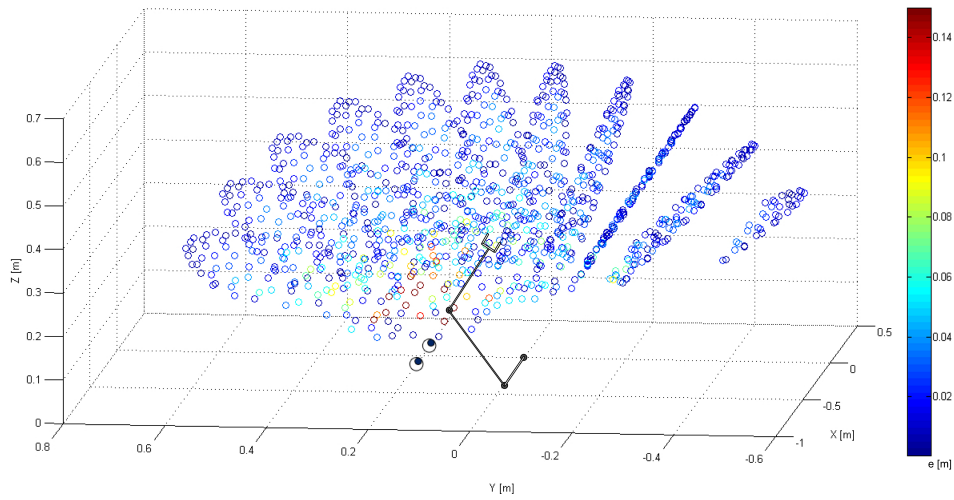


Figura 6.18: Nella figura viene illustrato l'errore euclideo fra la posizione effettiva dell'end effector e quella stimata dalla rete neurale; l'errore è abbastanza basso eccetto per quei punti nello spazio 3D che si trovano molto vicini alla testa e alla spalla.

Fig. 6.18). Quest'errore è dovuto ai vincoli dei giunti che abbiamo imposto, infatti il gomito si trova piegato a 132 gradi su 135. L'errore medio ottenuto può essere considerato un buon risultato, infatti la posizione dei giunti è calcolata prima che il braccio cominci a muoversi, esclusivamente a partire dall'input visivo, senza che venga utilizzato nessun sistema di retroazione, feedback tattile o stima della profondità basata sul movimento dell'oggetto (metodi alternativi per la stima della profondità citati in 2.2.5) .

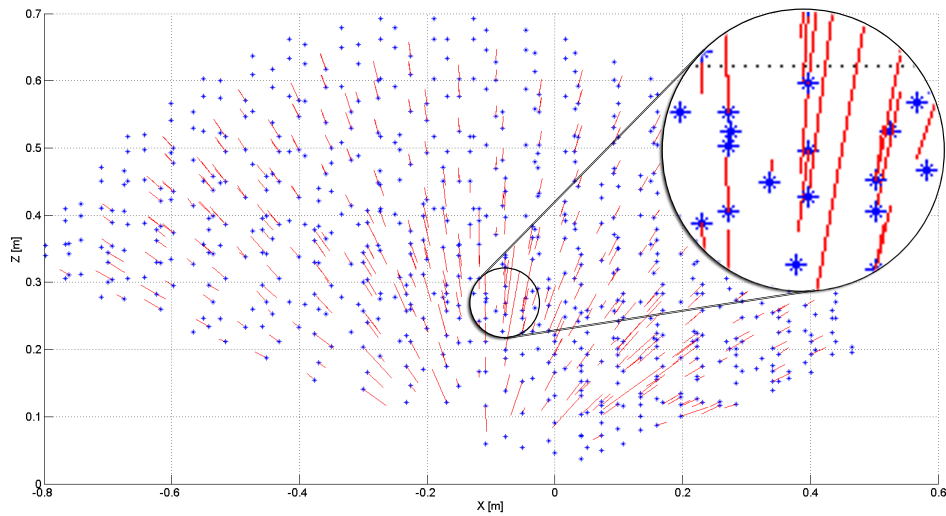


Figura 6.19: Direzione dell'errore nel piano X-Z. I punti blu rappresentano l'obiettivo, le linee rosse la distanza nello spazio fra l'obiettivo e la posizione del braccio calcolata dalla rete. Per mantenere l'immagine più leggibile, non tracciamo anche la posizione stimata dell'end effector.

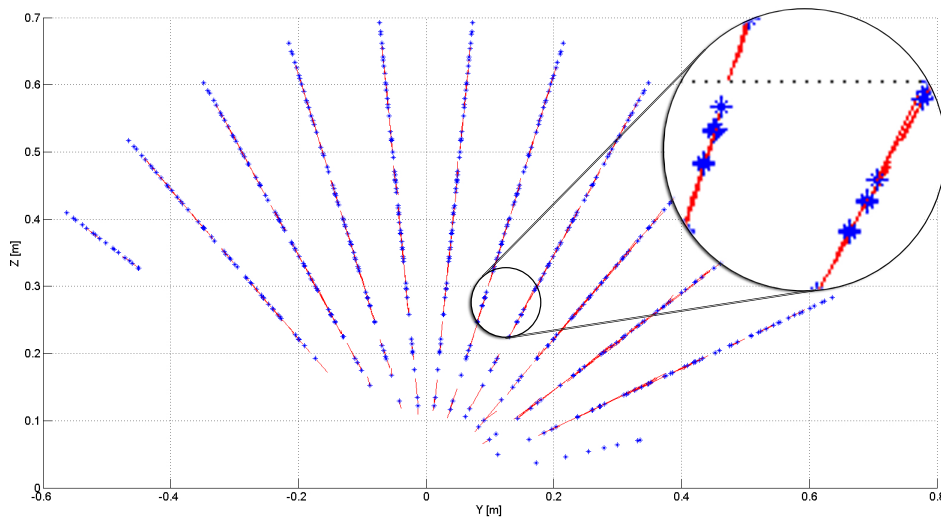


Figura 6.20: Direzione dell'errore nel piano Y-Z. I punti blu rappresentano l'obiettivo, le linee rosse la distanza nello spazio fra l'obiettivo e la posizione del braccio calcolata dalla rete. Per mantenere l'immagine più leggibile, non tracciamo anche la posizione stimata dell'end effector. Si osservi che i segmenti sono allineati lungo la stessa direzione, si veda Sez. 6.2.2 per ulteriori dettagli.

È interessante osservare la Figura 6.19: mostra la distanza euclidea fra la posizione stimata del braccio e l'obiettivo desiderato. I punti blu rappresentano l'obiettivo e i segmenti rossi la distanza nello spazio fra l'obiettivo e la posizione dell'end effector calcolata dalla rete. Il grafico è proiettato sul piano X-Z del sistema di riferimento del mondo, l'errore è distribuito su tutto lo spazio, ma decresce all'aumentare del valore di Z. Questo è dovuto ai limiti imposti al range di movimenti dei giunti del braccio. Nella Figura 6.20 è mostrato lo stesso grafico ma proiettato sul piano Y-Z; il fatto che tutti gli errori sono allineati evidenzia che essi non sono dovuti al nuovo grado di libertà (tilt) introdotto nella spalla, infatti le linee rosse sono allineate lungo la posizione della spalla stessa. Questo fenomeno è dovuto al fatto che la rete riceve in input solo una componente per il movimento di tilting della testa e da essa riesce a stimare facilmente la posizione del giunto spalla, mentre per gli altri due giunti deve stimare un movimento orizzontale a partire dalla composizione dei movimenti dei giunti degli occhi e del collo.

La Figura 6.21 mostra la distribuzione dei centri della RBF sullo spazio di input ed è riferita ad un singolo set di punti della Cross-folding validation. La fase di training ha selezionato circa 80 neuroni. I cerchi rossi rappresentano il centro di una funzione di base, i punti blu sono i dati di test e quelli azzurri i valori usati per la fase di addestramento. Nel pannello di sinistra è illustrata la distribuzione dei centri, che controlla il movimento di panning delle telecamere, proiettata sul piano θ_R - θ_L . I centri sono distribuiti lungo una linea, a riprova della grande cooperazione-competizione dei due angoli di panning. Inoltre, i valori di θ_R e θ_L usati come ingresso per la rete sono ben distribuiti lungo la stessa linea. Nel pannello centrale proiettiamo gli stessi centri sul piano θ_T - θ_N , osserviamo che la distribuzione dei centri è abbastanza uniforme prevalentemente perché la distribuzione dei due angoli è concettualmente indipendente. Infine, nel pannello destro proiettiamo i dati di input sul piano θ_L - θ_N e anche in questo caso abbia-

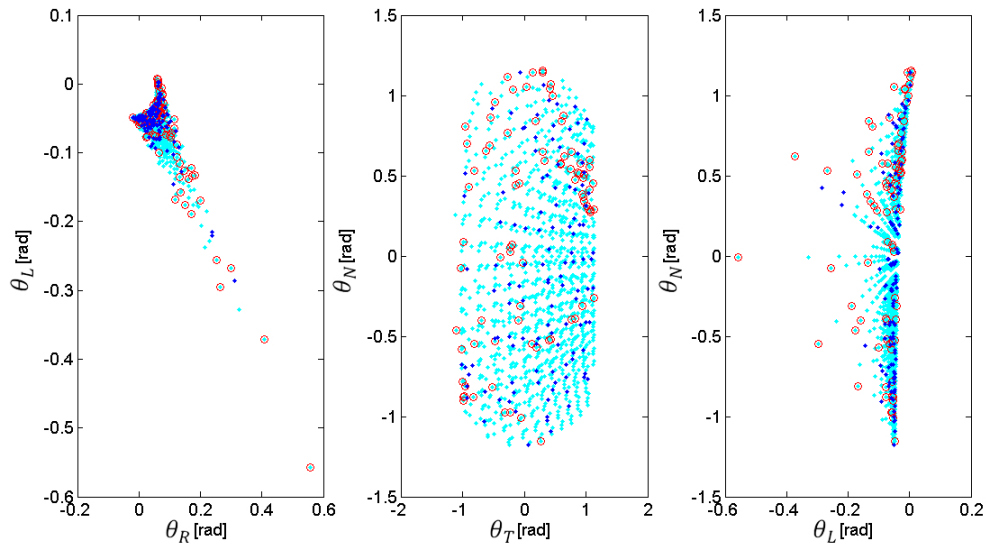


Figura 6.21: Posizionamento dei centri in corrispondenza dei valori in ingresso del sistema visivo. In azzurro i punti di training, in rosso i centri, in blu i valori casuali scelti per la simulazione.

mo una correlazione fra la componente del collo e la telecamera sinistra. È mostrato chiaramente che il collo esegue i movimenti grezzi, mentre la telecamera esegue solo piccoli movimenti per correggere la vergenza. Anche in questo caso i centri sono distribuiti prevalentemente nella regione contenente i movimenti comuni, ma la fase di training crea dei centri ad-hoc per gli “outliers”.

6.3 L'unione di Visione e Reaching

Avendo a disposizione il sistema di visione ed il sistema di reaching è ora possibile sfruttare la sinergia fra i due controllori per completare il nostro modello:

1. Il braccio robotico viene mosso selettivamente nello spazio.
2. Si calcola la cinematica diretta del braccio ottenendo il punto 3D raggiunto nello spazio visivo.
3. Il sistema di visione fovea l'end-effector.
4. La configurazione dei giunti del sistema di visione viene associata a quella del braccio robotico.
5. Si ripete da 1 a 4 più volte per generare il dataset che verrà utilizzato per addestrare la rete neurale.
6. Addestriamo e validiamo la rete neurale.

Prendendo ispirazione dal motor babbling, illustrato nella Sezione 2.4.1, la rete neurale modellizzata realizza il mapping visomotorio fra angoli del sistema di visione e di reaching, proponendo un'implementazione bioispirata per la trasformazione da sistema di riferimento visuale a sistema di riferimento motorio eseguito nel cervello umanoide dall'area V6A. Il controllore così realizzato permette quindi, dopo una fase di training, di calcolare in maniera istantanea le coordinate a cui portare i giunti del braccio per raggiungere un punto nello spazio. Non è più necessario che la feature da raggiungere venga foveata, infatti, una volta effettuato il mapping, il sistema è in grado di eseguire movimenti volontari del braccio verso la direzione indicata da pan e tilt del sistema visivo prima ancora che tali movimenti della testa siano effettuati (Fig. 6.22). I parametri di calibrazione della stereocamera e la propagazione degli errori di calcolo numerico nel controllore del

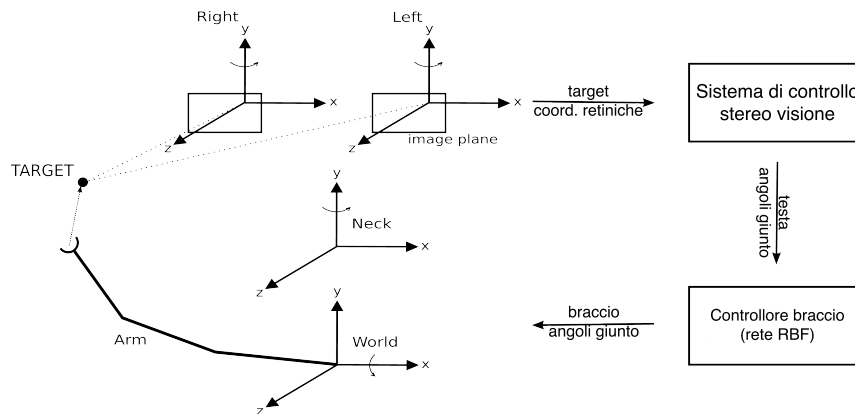


Figura 6.22: Sistema visomotorio finale.

sistema di visione, introducono un errore di 6 centimetri nella foveazione di un punto nello spazio. Quest'errore non si somma all'errore della rete neurale del sistema di reaching, essa è infatti addestrata utilizzando come target le coordinate dei giunti del braccio il cui punto raggiunto viene poi foveato e non viceversa. L'errore finale sarà quindi introdotto solamente dalla rete neurale, che trascurerà l'errore di foveazione.

Il sistema di mapping visomotorio per il braccio RR permette quindi di eseguire la trasformazione di sistemi di riferimento e compiere il raggiungimento dell'oggetto con un errore di 0.0047 metri. Per il braccio RRR l'errore medio nel reaching è di 0.0320 metri, a causa delle approssimazioni che introduce la rete (descritte in Sezione 6.2.2).

Capitolo 7

Conclusioni e sviluppi futuri

In questa tesi abbiamo sviluppato un nuovo sistema computazionale human-inspired capace di calcolare la mappa visomotoria fra un obiettivo (rilevato da un sistema di stereovisione attivo) ed un braccio a tre gradi di libertà. Il target è collocato nello spazio peripersonale 3D del robot modellizzato e viene percepito attraverso un sistema di visione costituito da due telecamere distanti l'una dall'altra 6 centimetri. Le due telecamere possono muoversi orizzontalmente (panning) e sono collocate su una testa collegata al resto del robot tramite un collo di 20 centimetri capace di effettuare movimenti di tilting e panning. Il sistema di stereovisione è ridondante e viene gestito tramite un controllore che implementa la legge di Hering dell'uguale innervazione. Il sistema di controllo realizzato permette alla struttura umanoide di occhi e collo, di comporre i movimenti per foveare un generico punto 3D nello spazio con traiettorie autovincolate in modo da portare la proiezione della feature nel centro del piano immagine, simulando ciò che accade nel processo umano di foveazione.

Il sistema di reaching è costituito da un braccio a tre gradi di libertà avente dimensioni medie simili a quelle umane: una spalla distante 20 centimetri dal collo, un braccio di 30 centimetri ed un avambraccio di 40 centimetri. Al termine di questa catena cinematica è posizionato l'end effector

che dovrà raggiungere il punto target. Il braccio realizzato è limitato nei movimenti da vincoli meccanici ispirati alle articolazioni umane, in modo da evitare posture innaturali.

Ispirandoci al motor babbling, il braccio viene mosso in modo casuale nello spazio visibile, permettendo al sistema di visione di foveare l'end effector. Sui dati generati viene addestrata la rete neurale RBF che permetterà successivamente di guidare il braccio in modo volontario. Tale rete calcola il mapping visomotorio fra angoli di foveazione del sistema visivo e angoli di rotazione dei giunti del braccio necessari a raggiungere il target foveato. La rete neurale implementata modella così il funzionamento dell'area V6A umana, che esegue la trasformazione da sistema di riferimento retinico a sistema di riferimento della mano.

Il nostro approccio è stato incrementale, dapprima abbiamo testato un sistema di base a tre gradi di libertà (angoli di panning degli occhi e tilting della testa), poi il sistema definito esteso dove i gradi di libertà introdotti sono quattro: panning del collo, oltre a quello degli occhi, e tilting della testa. Il controllore basato sulle leggi di Hering è stato scritto specificatamente per entrambi i sistemi gestendo la ridondanza del secondo caso.

Eseguendo il test di esplorazione dello spazio 3D e di reaching di una feature a partire da configurazioni diverse del sistema visivo si è evidenziato:

- un miglioramento prestazionale del sistema esteso. L'introduzione del panning nel collo permette infatti di foveare i punti nello spazio riducendo la varianza dell'errore lungo ogni direzione. L'errore risulta distribuito in maniera più uniforme, indipendente dalle coordinate del target e distribuito principalmente lungo la direzione che lo collega in linea d'aria con l'origine del sistema. Inoltre, la possibilità di ruotare il collo rende foveabili quei punti che sono visibili da solo uno dei due occhi quando il sistema di visione si trova nella sua

posizione iniziale.

- Una riduzione, oltre che dell'errore 2D sul piano immagine, anche dell'errore di foveazione nello spazio 3D. Inoltre tale errore si manifesta quasi esclusivamente con una percezione imprecisa (dell'ordine di pochi centimetri) sulla profondità visiva.
- Un aumento della robustezza del sistema che, a partire da una qualsiasi configurazione iniziale degli angoli, è in grado di foveare il target con un errore molto minore del sistema di base.
- Viene mantenuta una configurazione del sistema occhi-collo simile a quella umana.
- Le traiettorie generate dai movimenti di foveazione del target sono umanoidi: ad un rapido ed impreciso movimento di panning del collo e versione degli occhi segue un affinamento della foveazione tramite la vergenza.

Fornendo in ingresso alla rete neurale precedentemente addestrata gli angoli di foveazione del sistema visivo, si ottengono gli angoli dei giunti del braccio necessari a portare l'end effector in corrispondenza del target foveato. Il controllore globale è quindi in grado di rilevare un target nello spazio e a partire dagli angoli di foveazione, ancora prima di muovere la testa, calcolare gli angoli dei giunti necessari per raggiungere i punti con un errore medio di circa 3 centimetri. I risultati sono stati simulati e validati con una 10-fold Cross-validation che ha confermato la robustezza ed affidabilità del controllore.

In un approccio ispirato al developmental learning, il nostro sistema partendo da abilità "innate" quali la capacità di percepire una feature interessante nello spazio con il sistema visivo e l'applicazione ai giunti delle

forze necessarie per muovere un arto, è in grado di apprendere nuove competenze, di maggiore utilità e crescente complessità. Inizialmente è stato appreso come muovere gli occhi per foveare un punto nello spazio, raffinando successivamente questa conoscenza con l'aggiunta del movimento del collo. Poi, foveando il proprio end effector mosso casualmente, ha appreso come a determinati valori di pan e tilt possa associare determinati angoli di giunto. Su queste nuove abilità il sistema ha costruito la capacità di reaching guidata dalla vista; a partire da essa potrebbe essere implementati nuovi metodi per rendere il sistema ancora più preciso.

Sviluppi futuri

Il nostro lavoro di tesi rimane aperto a varie ed interessanti direzioni di sviluppo. Alcune sono legate all'implementazione fisica del sistema su di un robot reale, altre sono legate alla natura stessa dell'approccio incrementale utilizzato nel campo della ricerca. Il nostro controllore infatti può essere integrato in sistemi di complessità progressivamente crescente per svolgere innumerevoli attività, come ad esempio in un sistema più ampio volto a simulare l'elaborazione dell'intero processo visivo.

Focalizzandoci su un'area temporale a più breve termine, il prossimo passo consiste nel validare il modello su un sistema fisico, potendo raffinare l'apprendimento e correggere l'errore di reaching grazie all'aggiunta di un feedback tattile. Non è da trascurare la possibilità di sviluppare la cinematica del modello rendendola più complessa, avvicinandola quindi ulteriormente a quella umana e dotando così un eventuale robot umanoide della capacità di compiere movimenti più precisi ed evitare più agevolmente gli ostacoli.

Bibliografia

- [1] M. A. Goodale. Perception and action in the human visual system. In M. Editor Gazzaniga, editor, *The New Cognitive Neurosciences*, pages 365–377. The MIT Press, Cambridge, MA, 2000.
- [2] K. Vassie G. Morlino. Natural vs, via and pro artificial systems. *5th International Conference on Cognitive Systems*, und. rev.
- [3] M. Jeannerod, M. A. Arbib, G. Rizzolatti, and H. Sakata. Grasping objects: the cortical mechanisms of visuomotor transformation. *TINS*, 18(7):314–320, 1995.
- [4] A. Diamond. Developmental time course in human infants and infant monkeys, and the neural bases of inhibitory control in reaching. *Annals Of The New York Academy Of Sciences*, 608(d):637–676, 1990.
- [5] M. Marjanovic, B. Scassellati, and M. Williamson. Self-taught visually-guided pointing for a humanoid robot. *Camera*, pages 35–44, 1996.
- [6] B. Scassellati. How developmental psychology and robotics complement each other. In *Proceedings of the NSFDARPA Workshop on Development and Learning*. Michigan State University, 2000.
- [7] L. Berthouze and T. Ziemke. Epigenetic robotics: modelling cognitive development in robotic systems. *Cognitive Systems Research*, 6(3):189–192, 2005.

- [8] A. N. Meltzoff and M. K. Moore. Explaining facial imitation: a theoretical model. *Early development and parenting*, 6(34):179–192, 1997.
- [9] C. Von Hofsten. Eye-hand coordination in newborns. *Developmental European Journal of Developmental Psychology*, 18(3):450–461, 1982.
- [10] J. Piaget. *The origins of intelligence in children*. I.U. Press, New York, 1952.
- [11] J.G. Samarawickrama and S.P. Sabatini. Version and vergence control of a stereo camera head by fitting the movement into the hering’s law. In *Computer and Robot Vision, 2007. CRV '07. Fourth Canadian Conference on*, pages 363–370, 2007.
- [12] E. Chinellato, B. J. Grzyb, N. Marzocchi, A. Bosco, P. Fattori, and A. P. Del Pobil. The dorso-medial visual stream: From neural activation to sensorimotor interaction. *Neurocomputing*, 74:1203–1212, 2011.
- [13] M. J. L. Orr. Introduction to radial basis function networks. *Time*, 302(5652):1–67, 1996.
- [14] D. Kriesel. A brief introduction to neural networks. pages 105–120, 2007.
- [15] R. Rojas. *Neural networks: a systematic introduction*, pages 11–26. Springer, Berlin, 1996.
- [16] R. Hartley and A. Zisserman. *Multiple View Geometry in computer vision*. Cambridge University press, Cambridge, 2003.
- [17] G. Gini and V. Caglioti. *Robotica*. Edizioni Zanichelli, Bologna, 2003.
- [18] M. W. Spong, S. Hutchinson, and M. Vidyasagar. *Robot Modeling and Control*. J. Wiley and Sons, New York, 2005.

- [19] A. K. E. Horn and R. J. Leigh. The anatomy and physiology of the ocular motor system. *Handbook of clinical neurology*, 102:21–69, 2011.
- [20] F. Crick. Function of the thalamic reticular complex: the searchlight hypothesis. *PNAS*, 81(14):4586–4590, 1984.
- [21] E. Chinellato, M. Antonelli, B. J. Grzyb, and A. P. Del Pobil. Implicit sensorimotor mapping of the peripersonal space by gazing and reaching. *IEEE Transactions on Autonomous Mental Development*, 3(1):43–53, 2011.
- [22] M. Hoffmann, H. G. Marques, A. H. Arieta, H. Sumioka, M. Lungarella, and R. Pfeifer. Body schema in robotics: A review. *IEEE Transactions on Autonomous Mental Development*, 2(4):304–324, 2010.
- [23] R. Saegusa, G. Metta, G. Sandini, and S. Sakka. Active motor babbling for sensorimotor learning. *2008 IEEE International Conference on Robotics and Biomimetics*, pages 794–799, 2009.
- [24] S. Chen, Y. Li, and N. M. Kwok. Active vision in robotic systems: A survey of recent developments. *The International Journal of Robotics Research*, 30(11):1343–1377, 2011.
- [25] Y. Wang and B. E. Shi. Autonomous development of vergence control driven by disparity energy neuron populations. *Neural Computation*, 22(3):730–751, 2010.
- [26] Y. Wang and B. E. Shi. Improved binocular vergence control via a neural network that maximizes an internally defined reward. *IEEE Transactions on Autonomous Mental Development*, 3(3):247–256, 2011.
- [27] A. Gibaldi, M. Chessa, A. Canessa, S. P. Sabatini, and F. Solari. A cortical model for binocular vergence control without explicit calcula-

- tion of disparity. *Proc European Symposium on Artificial Neural Networks ESANN09*, 73(7-9):1065–1073, 2010.
- [28] A. Gibaldi, A. Canessa, M. Chessa, S.P. Sabatini, and F. Solari. A neuro-morphic control module for real-time vergence eye movements on the icub robot head. In *Humanoid Robots (Humanoids), 2011 11th IEEE-RAS International Conference on*, pages 543–550, 2011.
- [29] K. Shimonomura and T. Yagi. Neuromorphic vergence eye movement control of binocular robot vision. In *Robotics and Biomimetics (ROBIO), 2010 IEEE International Conference on*, pages 1774–1779, 2010.
- [30] R. Saegusa, G. Metta, and G. Sandini. Own body perception based on visuomotor correlation. *IEEE/RSJ Int Intelligent Robots and Systems (IROS)*, 2:1044–1051, 2010.
- [31] Y. Burnod, P. Baraduc, A. Battaglia-Mayer, E. Guigon, E. Koechlin, S. Ferraina, F. Lacquaniti, and R. Caminiti. Parieto-frontal coding of reaching: an integrated framework. *Experimental brain research*, 129(3):325–346, 1999.
- [32] T. M. Martinetz, H. J. Ritter, and K. J. Schulten. Three-dimensional neural net for learning visuomotor coordination of a robot arm. *Neural Networks IEEE Transactions on*, 1(1):131–136, 1990.
- [33] M. Hersch, E. Sauser, and A. Billard. Online learning of the body schema. *International Journal of Humanoid Robotics*, 5(2):161–181, 2008.
- [34] C. Goerick, H. Wersing, I. Mikhailova, and M. Dunn. Peripersonal space and object recognition for humanoids. *5th IEEE/RAS International Conference on Humanoid Robots 2005*, pages 387–392, 2005.

- [35] L. L. Chen. Head movements evoked by electrical stimulation in the frontal eye field of the monkey: evidence for independent eye and head control. *Journal of Neurophysiology*, 95(6):3528–3542, 2006.

In collaborazione con:

Airlab PoliMI



Si ringrazia il
Laboratorio di Intelligenza Artificiale,
Dipartimento di Informatica,
Università di Zurigo

realizzato con:

L^AT_EX

MatlabTM

elaborazioni eseguite su server

Nvidia CUDA

