

ALMA MATER STUDIORUM - UNIVERSITA' DI BOLOGNA
CAMPUS DI CESENA
SCUOLA DI INGEGNERIA E ARCHITETTURA

CORSO DI LAUREA MAGISTRALE IN INGEGNERIA BIOMEDICA

**Studio e applicazione di tecniche di riduzione di
dimensionalità a modelli geometrici di percezione visiva**

Tesi in:
Bioimmagini e visione LM

Relatore:

Prof. CLAUDIO LAMBERTI

Correlatori:

Prof. ALESSANDRO SARTI

Dott. GIACOMO COCCI

Presentata da:

MARTA FAVALI

SESSIONE II

ANNO ACCADEMICO 2012–2013

Indice

Introduzione	ii
1 Caratteristiche funzionali del sistema visivo	1
1.1 L'occhio e il percorso visivo primario	1
1.2 I profili recettori	11
1.2.1 I filtri di Gabor	18
1.3 Organizzazione della corteccia visiva	21
2 La percezione visiva:	
dalle teorie della Gestalt al modello differenziale Citti-Sarti	26
2.1 I campi di associazione	29
2.2 Modello differenziale Citti-Sarti per il completamento percettivo	35
3 Connettività corticale tra neuroni	44
3.1 Connessioni orizzontali in corteccia	44
3.2 Costruzione di kernel stocastici	52
4 Applicazione del modello di connettività:	
 il raggruppamento percettivo	62
4.1 Grouping e clustering.	63
4.2 Spectral Clustering.	65
4.3 Matrice di affinità.	75
4.4 Esempi applicativi.	87
4.4.1 Articolazione figura-sfondo.	87
4.4.2 Estrazione di figure attraverso metodi di riduzione di dimensionalità	97
Conclusioni	108

Bibliografia **111**

Ringraziamenti **116**

Introduzione

La tesi affronta il tema della neuromatematica della visione, in particolare l'integrazione di modelli geometrici di percezione visiva con tecniche di riduzione di dimensionalità.

Dall'inizio del secolo scorso, la corrente ideologica della Gestalt iniziò a definire delle regole secondo le quali stimoli visivi distinti tra loro possono essere percepiti come un'unica unità percettiva, come ad esempio i principi di prossimità, somiglianza o buona continuazione.

Nel tentativo di quantificare ciò che gli psicologi avevano definito in maniera qualitativa, Field, Hayes e Hess hanno descritto, attraverso esperimenti psicofisiologici, dei *campi di associazione* per stimoli orientati, che definiscono quali caratteristiche due segmenti dovrebbero avere per poter essere associati allo stesso gruppo percettivo.

Grazie alle moderne tecniche di neuroimaging che consentono una mappatura funzionale dettagliata della corteccia visiva, è possibile giustificare su basi neurofisiologiche questi fenomeni percettivi. Ad esempio è stato osservato come neuroni sensibili ad una determinata orientazione siano preferenzialmente connessi con neuroni aventi selettività in posizione e orientazione coerenti con le regole di prossimità e buona continuazione.

Partendo dal modello di campi di associazione nello spazio $\mathbb{R}^2 \times S^1$ introdotto da Citti e Sarti, che introduce una giustificazione del completamento percettivo sulla base della funzionalità della corteccia visiva primaria (V1), è stato possibile modellare la connettività cellulare risolvendo un sistema di equazioni differenziali stocastiche. In questo modo si sono ottenute delle densità di probabilità che sono state interpretate come probabilità di connessione tra cellule semplici in V1.

A queste densità di probabilità è possibile collegare direttamente il concetto di *affinità* tra stimoli visivi, e proprio sulla costruzione di determinate matrici di affinità si sono basati diversi metodi di riduzione di

dimensionalità. La fenomenologia del grouping visivo descritta poco sopra è, di fatto, il risultato di un procedimento di riduzione di dimensionalità. I risultati ottenuti da questa analisi e gli esempi applicativi sviluppati si sono rivelati utili per comprendere più nel dettaglio la possibilità di poter riprodurre, attraverso l'analisi spettrale di matrici di affinità calcolate utilizzando i modelli geometrici di Citti-Sarti, il fenomeno percettivo di grouping nello spazio $\mathbb{R}^2 \times S^1$.

“Il tutto è più della somma delle singole parti”

Capitolo 1

Caratteristiche funzionali del sistema visivo

L'occhio, o bulbo oculare, è l'organo di senso che permette il fenomeno della visione: ha il compito di ricavare informazioni sull'ambiente circostante attraverso la luce. L'occhio umano (e degli organismi superiori) raccoglie la luce che gli proviene dall'ambiente, ne regola l'intensità attraverso un diaframma (l'iride), la focalizza attraverso un sistema regolabile di lenti per formarne un'immagine che trasforma in una serie di segnali elettrici che attraverso il nervo ottico vengono inviati al cervello per l'elaborazione e l'interpretazione, raggiungendo il nucleo genicolato laterale (LGN). Gli assoni dei neuroni talamici raggiungono poi la parte occipitale del cervello, dove la corteccia visiva primaria (nota anche come corteccia striata o semplicemente V1) inizia a processare le informazioni attraverso i suoi diversi strati. In questo capitolo, dopo una prima descrizione anatomica e fisiologica dell'occhio, sono introdotti i profili recettori e i filtri di Gabor, come filtri che descrivono i profili di cellule in V1. Verrà poi illustrata l'organizzazione della corteccia visiva e la sua struttura ipercolonnare.

1.1 L'occhio e il percorso visivo primario

Gli occhi umani sono posizionati nelle due cavità orbitarie del cranio. La superficie esterna di ciascun occhio è una sfera incompleta e opaca costituita per il 93% dalla sclera, che anteriormente si interrompe dando origine a una seconda porzione sferica trasparente più piccola, per circa il 7% della superficie totale e di differente curvatura, la cornea. Internamente alla sclera è presente uno strato pigmentato e vascolarizzato, la coroide, che continua anteriormente dando origine prima al corpo ciliare e poi all'iride, un disco membranoso con un foro centrale, la pupilla, il cui diametro è

regolato dalla muscolatura dell'iride. In fondo alla coroide si estende uno strato fotosensibile, la retina, che termina anteriormente a livello dell'ora serrata, la giunzione tra la coroide stessa e il corpo ciliare. La retina contiene i fotorecettori dell'occhio che trasducono la luce in potenziali elettrici inviati attraverso il nervo ottico. Dal corpo ciliare si dipartono le fibre che si attaccano al cristallino, la lente dell'occhio, collocata posteriormente all'iride e a livello del corpo ciliare [15].

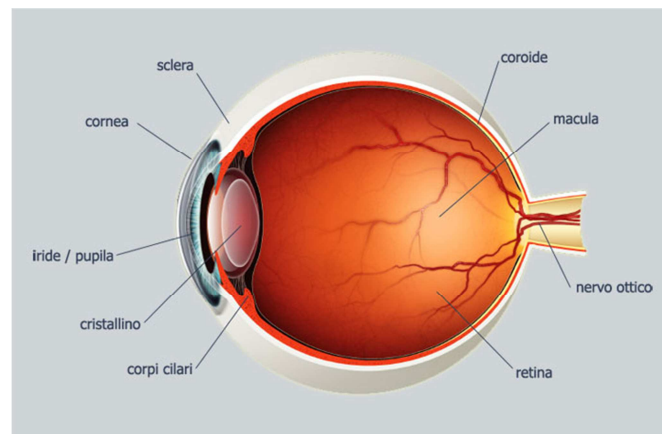


Figura 1.1: Anatomia dell'occhio.

Esaminando l'anatomia del sistema visivo umano si possono distinguere tre unità principali: la retina, il nucleo genicolato laterale e la corteccia visiva. L'insieme di questi tre blocchi costituisce il percorso ottico principale e, attraverso questo percorso, le informazioni provenienti dal mondo esterno sono elaborate fino a giungere alle aree corticali associative dove vengono integrate con le informazioni provenienti da altri canali sensitivi.

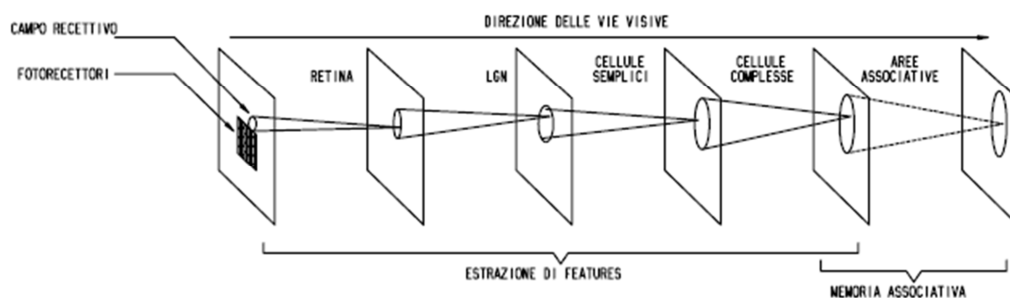


Figura 1.2: Schema del percorso visivo primario.

La luce passa l'intero spessore della retina e colpisce i fotorecettori. Questi sono principalmente di due tipi. I coni, più corti, localizzati quasi esclusivamente nella parte centrale della retina (area maculare), sono specializzati alla visione in condizioni di alta luminosità e traggono contatti diretti con uno strato di cellule neuronali deputate alla trasmissione dell'impulso visivo. Si viene a creare, quindi, una trasmissione di uno a uno cioè, un cono scarica il suo impulso o per meglio dire, trae contatto con una sola cellula nervosa. Tale forma di collegamento è altamente strutturato e fa sì che i coni siano deputati alla visione diurna, a quella dei colori e del contrasto.

I bastoncelli, di forma più allungata, sono molto più numerosi dei coni e risiedono maggiormente nella parte periferica della retina. Essi sono specializzati a raccogliere stimoli luminosi di bassa intensità e quindi sfruttati nella visione in condizioni di scarsa luminosità [15].

La retina, la cui funzione è simile ad una pellicola fotografica, riveste la superficie interna del globo oculare. E' una sottile membrana trasparente formata da due tipi di cellule e suddivisa in due aree: un'area centrale chiamata macula che contiene la fovea, ricca di coni; un'area media e periferica che serve a mediare la visione crepuscolare e notturna, dove prevalgono le cellule dei bastoncelli.

Dopo aver attraversato la cornea, la camera anteriore, la pupilla, il cristallino ed il vitreo, i raggi luminosi provenienti dal punto di fissazione vengono fatti convergere sulla retina ed in particolare in quella piccolissima area chiamata fovea: una struttura altamente specializzata che presiede, alla massima acuità visiva per lontano e vicino, alla percezione dei colori e alla sensibilità al contrasto. Nella retina iniziano i meccanismi più complessi della visione.

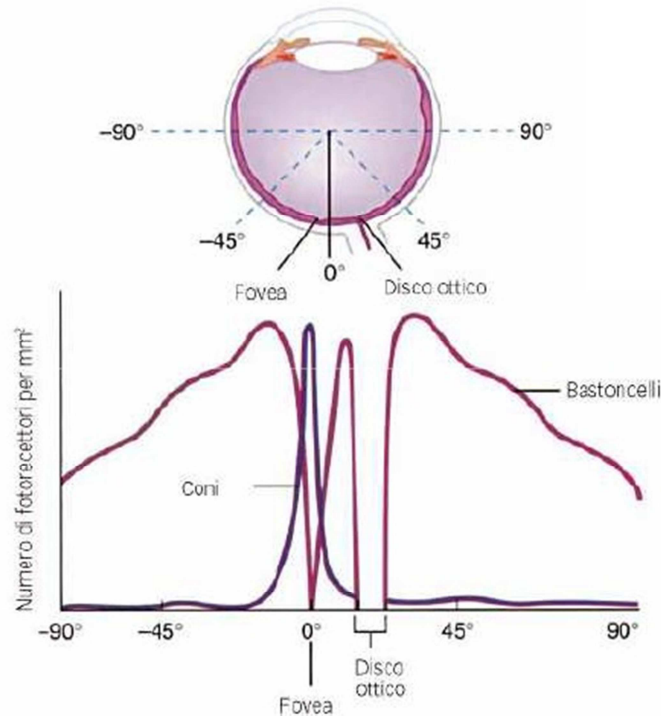


Figura 1.3: Distribuzione di coni e bastoncelli nella retina.

I fotorecettori della retina convertono quindi le intensità luminose della scena in segnali elettrochimici, dando così una prima rappresentazione dell'immagine. La retina, tuttavia, non si limita a registrare passivamente le immagini formate sulla sua superficie, ma le scompone analizzandone i molteplici parametri fisici. Nella retina, infatti, oltre ai fotorecettori, si possono individuare altre quattro principali classi di interneuroni: i) le cellule orizzontali, che trasmettono orizzontalmente i segnali ricevuti dai fotorecettori; (ii) le cellule bipolari, che inviano i segnali alle cellule gangliari; (iii) le cellule amacrine, che hanno funzioni simili alle orizzontali, ma operano in uno strato sottostante; (iv) le cellule gangliari, che con le loro terminazioni assoniche trasmettono il messaggio nervoso ai centri superiori, sono le prime cellule del sistema visivo nelle quali si generano potenziali d'azione.

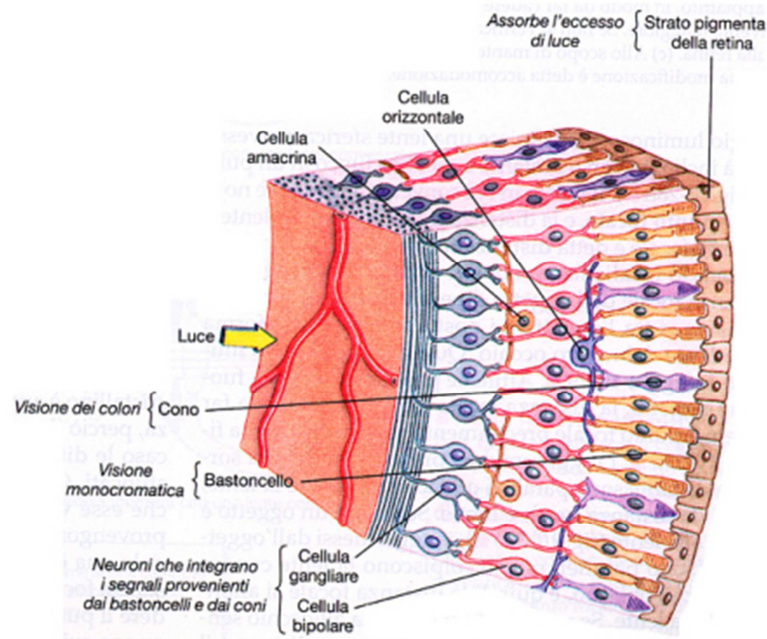


Figura 1.4: Struttura della retina.

Le cellule gangliari possono essere distinte in cellule M (per magnae o grandi) e cellule P (per parvae o piccole). Le cellule M sono connesse con un gran numero di coni e bastoncelli e per questo sono in grado di fornire informazioni sul movimento di un oggetto o su rapide variazioni del contesto luminoso. Le cellule P sono connesse con un minor numero di recettori e sono adatte a fornire informazioni sulla forma e sul colore di un oggetto. Questa distinzione anatomo/funzionale è importantissima: le informazioni visive provenienti dalle cellule M e dalle cellule P rimangono segregate anche a livello centrale. Queste informazioni si collegano a zone distinte del nucleo genicolato laterale; inoltre, quest'ultimo le informazioni provenienti dai 2 tipi cellulari, vengono smistate a zone diverse della corteccia striata. Da qui, le informazioni prendono due vie distinte; la via ventrale, per le informazioni in gran parte provenienti dalle cellule P e la via dorsale, per le informazioni in gran parte provenienti dalle cellule M [19].

L'informazione sensoriale così elaborata viene codificata dalle cellule gangliari e trasmessa lungo gli assoni di queste principalmente al nucleo

genicolato laterale. Il nucleo genicolato laterale è una stazione talamica preposta allo smistamento delle informazioni visive provenienti dal tratto ottico, una continuazione del nervo ottico (composto dagli assoni delle cellule gangliari che risiedono nella retina nasale e nella retina temporale di un singolo occhio). Il tratto ottico di destra è composto dalle fibre nervose provenienti dall'emiretina temporale dell'occhio destro e dall'emiretina nasale dell'occhio sinistro; viceversa, il tratto ottico di sinistra è composto dalle fibre provenienti dall'emiretina temporale dell'occhio sinistro e dall'emiretina nasale dell'occhio destro. In pratica, ciascun tratto ottico contiene la rappresentazione completa dell'emicampo visivo controlaterale.

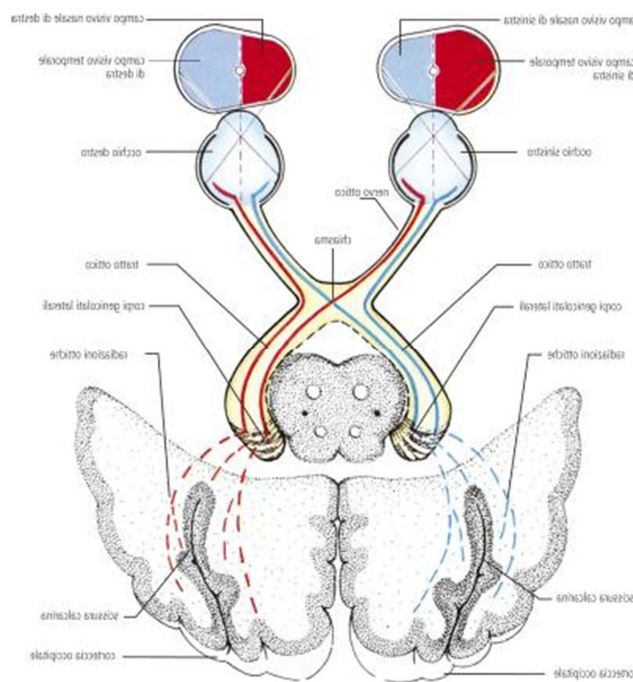


Figura 1.5: Separazione delle due parti del campo visivo.

I segnali elettrici, che attraverso il nervo ottico vengono inviati al cervello per essere elaborati, raggiungono il nucleo genicolato laterale (LGN). Gli assoni dei neuroni talamici raggiungono poi la parte occipitale del cervello, dove la corteccia visiva primaria inizia a processare le informazioni attraverso i suoi diversi strati.

Si avrà, dunque, che la corteccia visiva destra "vedrà" il campo visivo di sinistra, mentre la corteccia visiva sinistra "vedrà" il campo visivo di destra. Il tratto ottico decorre dal chiasma ottico (nel quale si ha quindi un parziale incrociamiento delle fibre nervose provenienti dalle due emiretine di ciascun occhio) al nucleo genicolato laterale del talamo.

Attraverso il nucleo genicolato laterale (LGN) passano quasi tutti gli input sensoriali verso la corteccia, ad eccezione di quelli olfattivi. Nella proiezione dalla retina al LGN vengono preservate le caratteristiche topografiche dello stimolo (ordine retinotopico), infatti gli assoni provenienti dalla retina vanno a connettersi con i neuroni del LGN secondo un ordine preciso che dipende dalla posizione dei relativi neuroni sulla retina.

Il nucleo genicolato laterale è diviso in 6 lamine concentriche; due di queste hanno afferenze esclusive dalle cellule gangliari M, mentre le restanti 4 ricevono da cellule gangliari P. L'analisi in parallelo delle informazioni veicolate dalle cellule gangliari P ed M è particolarmente importante per la corretta percezione del colore e dell'illuminazione, oltre a fornire canali preferenziali per i processi di discriminazione spaziale e temporale [15]. Questa stazione sembra svolgere una semplice azione di trasferimento, ma per le sue dimensioni e per la posizione che occupa, potrebbe rivestire un ruolo ben più importante. In particolare, l'esistenza di un certo numero di sinapsi inibitrici e la consistente presenza di terminazioni di retroazione che dalla corteccia si innervano nel talamo, fa supporre che il LGN abbia una funzione importante nell'integrazione spazio-temporale dell'informazione. Dai neuroni del LGN partono le fibre nervose dirette alla corteccia cerebrale, dove l'elaborazione dei segnali visivi viene perfezionata rendendo possibile la percezione cosciente. La corteccia visiva è suddivisa in un certo numero di aree che possono essere "organizzate" gerarchicamente e ciascuna di esse presenta una mappa retinotopica più o meno precisa. La prima area della corteccia visiva (V1) è la cosiddetta area 17 (secondo la classificazione di Brodmann), detta anche area striata per il suo particolare

aspetto (dovuto a fasci di fibre afferenti); qui si innervano le terminazioni provenienti dal LGN e viene rappresentato in modo abbastanza dettagliato il campo visivo. Ciascuna porzione di quest'area risponde quindi prevalentemente a proprietà locali di una piccola zona del campo visivo. Considerando aree a livello più alto nella gerarchia, il "mappaggio" diventa meno definito e nello stesso tempo i neuroni rispondono a caratteristiche sempre più complesse dello stimolo visivo. Aree differenti sono inoltre specializzate nell'estrarre caratteristiche differenti, come un'area sensibile prevalentemente al movimento (V5), un'altra al colore (V4) e così via.

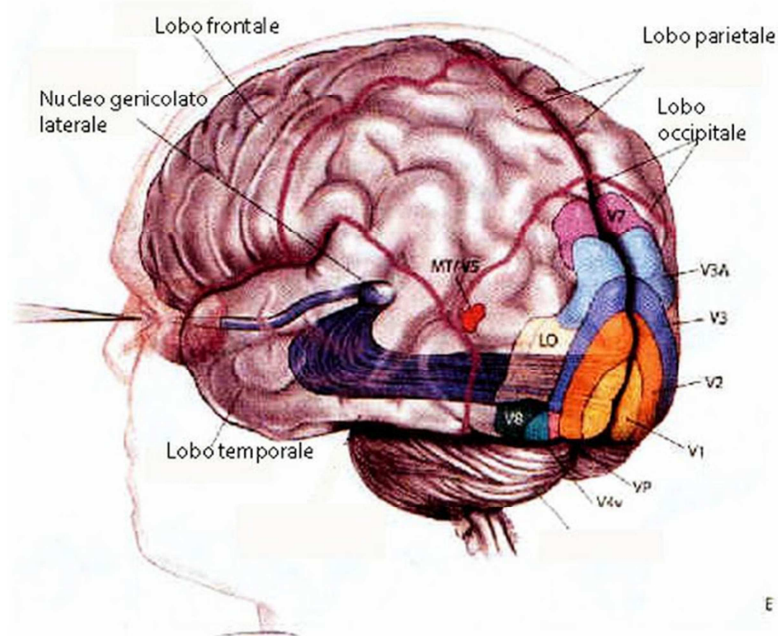


Figura 1.6: Aree della corteccia visiva.

Nello specifico l'area visiva V2, nota anche come corteccia pre-striata, è la seconda area principale nella corteccia visiva, e la prima regione all'interno dell'area associativa visiva. Riceve forti connessioni di feedforward dall'area V1 e invia forti connessioni alla V3, V4 e V5 oltre a forti connessioni di retroregolazione alla V1.

Anatomicamente, la V2 è divisa in quattro quadranti, con una rappresentazione dorsale e ventrale negli emisferi. Assieme, queste quattro regioni forniscono una mappa completa del mondo visivo. Funzionalmente, la V2 ha molte proprietà in comune con la V1. Le cellule sono "sintonizzate" per caratteristiche semplici come l'orientazione, la frequenza spaziale e il colore. Le risposte di molti neuroni V2 sono anche modulate da proprietà più complesse, come l'orientazione di contorni illusori e del fatto che lo stimolo faccia parte della figura oppure dello sfondo. Inoltre l'area visiva V5, detta anche Area visiva MT (medio temporale), è una regione della corteccia visiva extrastriata la quale si suppone giochi un ruolo fondamentale nella percezione del movimento, nell'integrazione di segnali locali di movimento in percetti globali e nella guida di alcuni movimenti oculari.

La corteccia visiva primaria è suddivisibile in almeno sei strati, sulla base della disposizione delle cellule su piani paralleli alla superficie corticale. Le afferenze dal LGN terminano soprattutto nel quarto strato, mentre gli strati 2 e 6 producono la maggior parte dei segnali di uscita. In V1 la maggior parte delle fibre proveniente dal LGN terminano nello strato 4 che è suddiviso in tre sottostrati: A, B, C. Il C è ulteriormente suddiviso in $4C\alpha$ e $4C\beta$. Le proiezioni degli strati parvocellulari del LGN terminano nei substrati $4A$, $4C\beta$ e nella parte superiore dello strato 6. Le cellule nello strato $4C\beta$ proiettano ai neuroni degli strati 2, e 3. Gli strati magnocellulari terminano in $4C\alpha$ e nella parte inferiore dello strato 6. Le cellule di $4C\alpha$ proiettano allo strato 4B (figura 1.7). Entrambe le M e P mandano segnali alla corteccia visiva.

I segnali provenienti dai due occhi sono mantenuti separati nello strato 4, infatti le cellule di uno strato del LGN proiettano a un gruppo di cellule bersaglio nello strato 4C, separato dai gruppi a cui giungono i segnali dall'altro occhio [15].

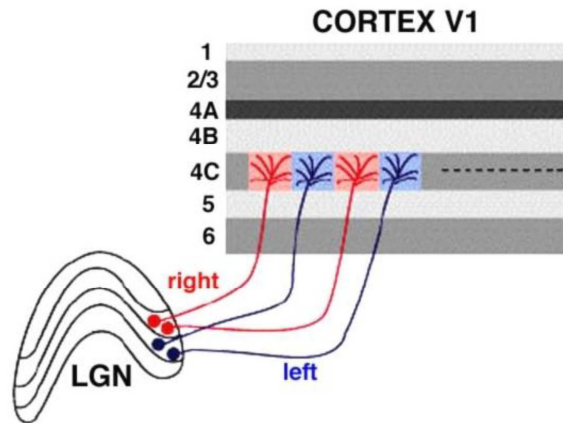


Figura 1.7:

La figura mostra come i segnali provenienti dai due occhi siano mantenuti separati nello strato 4.

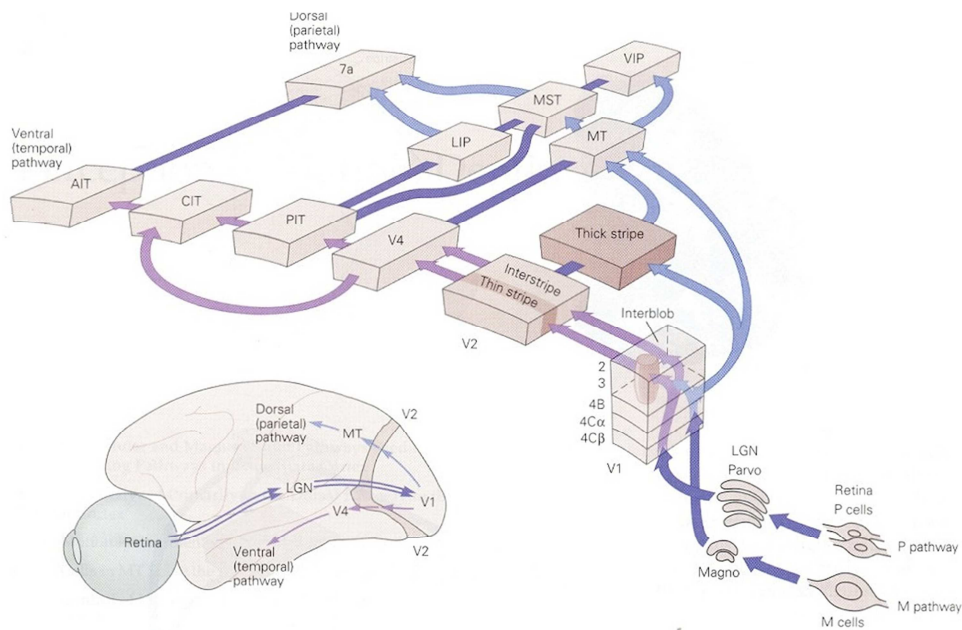


Figura 1.8: Diagramma funzionale del percorso visivo.

Al di sotto e al di sopra di questo strato, la maggior parte delle cellule riceve informazioni da entrambi gli occhi, sebbene un occhio sia generalmente dominante. Hubel e Wiesel hanno nominato queste colonne di cellule colonne di dominanza oculare.

Prima però di spiegare completamente l'organizzazione funzionale della corteccia visiva occorre introdurre il concetto di campo recettore.

1.2 I profili recettori

Il campo recettore è una caratteristica di tutte le cellule nervose sensoriali ed in quelle appartenenti al percorso visivo indica la regione dell'area visiva in cui deve presentarsi lo stimolo per eccitare o inibire il neurone interessato. Ad eccezione dei fotorecettori e degli elementi delle prime stazioni retiniche, che rispondono semplicemente alla presenza o assenza di luce, la risposta dei neuroni non è in genere uniforme e la luce diffusa a tutto il campo recettore rappresenta uno stimolo assolutamente inefficace. Generalmente, quindi, il profilo recettore non è omogeneo, ma è caratterizzato da sottoregioni denominate di tipo “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 (depolarizzazione), mentre uno stimolo scuro ne provoca l'iperpolarizzazione (inibizione dell'attività elettrica). Del tutto complementare è il comportamento del neurone a seguito della stimolazione di una sottoregione “OFF” del suo profilo recettore: l'assenza di uno stimolo luminoso accresce la sua attività spontanea, mentre la sua presenza ne inibisce la risposta. Considerando i neuroni come sistemi lineari tempo varianti, i profili recettori possono essere visti come la risposta all'impulso di questo sistema e a partire da quest'ultima è possibile calcolare l'Output del sistema attraverso una sua convoluzione con l'Input. Il profilo recettore agisce come un filtro sull'immagine tenendo conto di queste considerazioni, un profilo recettore può essere modellato da una funzione bidimensionale definita sul piano immagine \mathbb{R}^2 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 [8].

Le cellule appartenenti al primo strato del sistema visivo sono eccitate in modo diretto dagli stimoli luminosi e i loro profili recettori sono determinati esclusivamente dal modo in cui tali cellule interagiscono tra loro sullo stesso strato. Le cellule degli strati successivi, invece, sono eccitate dagli stimoli luminosi solo per via indiretta: lo stimolo viene alterato da elaborazioni compiute dagli strati precedenti. Dunque i profili recettori di cellule appartenenti a strati successivi al primo, sono determinati sia dalle elaborazioni degli strati precedenti, sia dal modo in cui esse interagiscono tra loro; se le operazioni compiute sullo stimolo sono lineari, si può parlare di sovrapposizione degli effetti [15]. Salendo a livelli superiori dell'apparato visivo, i profili recettori si sovrappongono, le loro dimensioni aumentano e lo stimolo efficace diviene progressivamente più complesso (da stimoli puntiformi a barre luminose orientate fino a tessiture vere e proprie). Un primo esempio viene dalle cellule gangliari della retina (figura 1.9), i cui profili recettori sono in linea di massima più ampi di quelli dei fotorecettori e in parte sono sovrapposti. Una tipica cellula gangliare risponde molto debolmente ad un'illuminazione uniforme. Inoltre, una piccola chiazza luminosa che occupi solo una parte del profilo recettore della cellula esercita effetti diametralmente opposti a seconda che sia situata al centro o alla periferia del campo stesso: ad esempio, può stimolare l'attività neurale se si trova al centro, ma inibirla se si trova in posizione periferica. Per una cellula gangliare di questo tipo (ON-OFF), lo stimolo più efficace è rappresentato da un disco luminoso circondato da un alone circolare scuro. Altre cellule gangliari (OFF-ON) rispondono più intensamente ad un modello luminoso inverso: un disco scuro con un alone luminoso [8].

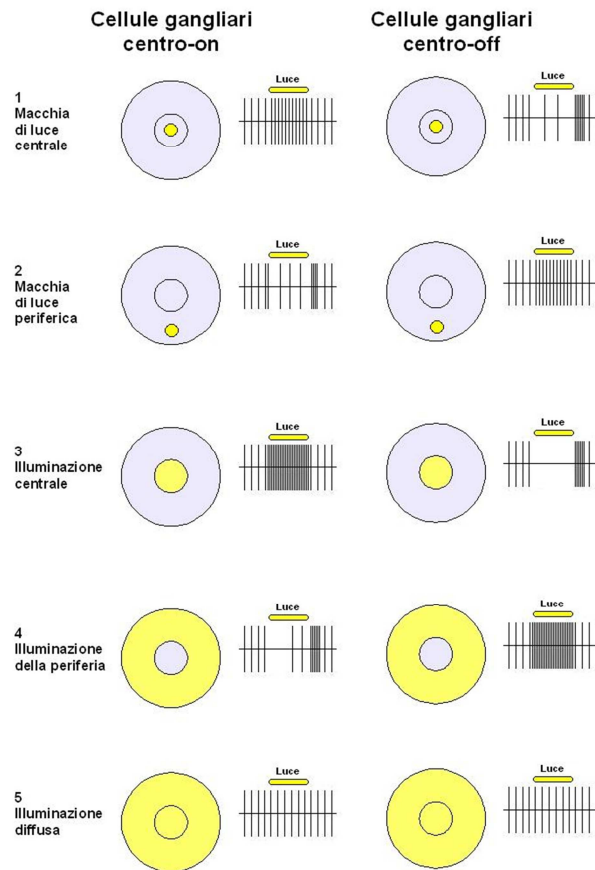


Figura 1.9:

Numero di potenziali di azione al secondo (posti a destra dei profili recettori) prodotti per uno stesso tipo di stimolo luminoso che investe il profilo recettore di una cellula gangliare ON center e OFF center.

Le cellule situate a livello della corteccia possiedono requisiti ancora più specifici. Una numerosa classe di neuroni corticali risponde solamente ad una barra luminosa o scura posta su uno sfondo di luminosità contraria; per un dato neurone la barra deve essere posta in una regione particolare del campo visivo e possedere una particolare orientazione rispetto all'asse orizzontale. Altri neuroni corticali per rispondere richiedono non solo che lo stimolo abbia una particolare orientazione, ma anche una particolare frequenza spaziale oppure che si muova in una determinata direzione o che possieda dimensioni peculiari. Si riportano ora le caratteristiche dei profili recettori delle tre unità del sistema visivo.

Per quanto riguarda la retina l'analisi si riduce alle cellule gangliari. I loro profili possono essere descritti da una funzione a forma circolare, che li rende insensibili all'orientazione dello stimolo e si comportano dunque come semplici rivelatori di contrasto.

Nel nucleo genicolato laterale si trovano profili recettori molto simili nella forma a quelli delle cellule gangliari ma sono più estesi, di cui condividono, oltre la forma circolare, la struttura concentrica delle sottoregioni ON e OFF (figura 1.11a). E' possibile modellare questi profili recettori attraverso una funzione a "cappello messicano", cioè un laplaciano di G, dove G è una campana di Gauss di larghezza σ :

$$(1.1) \quad G(x, y, \sigma) = e^{-\frac{(x^2 + y^2)}{2\sigma^2}}$$

mentre il suo laplaciano viene scritto come:

$$(1.2) \quad \Delta G = \frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2}$$

La larghezza σ della funzione gaussiana rappresenta la dimensione della campana: quanto più grande è σ maggiore sarà l'estensione della gaussiana e quindi la dimensione del profilo recettore.

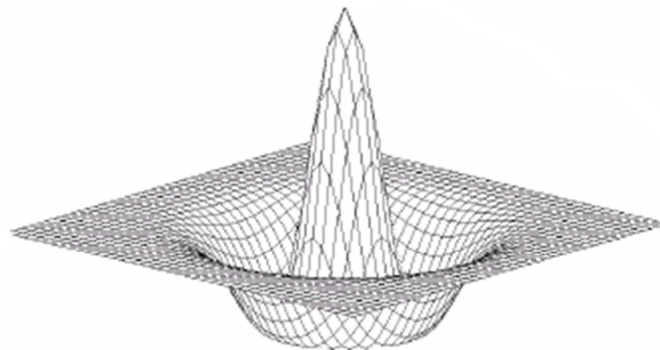


Figura 1.10:

Rappresentazione di un laplaciano di gaussiana, come modello per i profili recettori del nucleo genicolato laterale.

Per quanto riguarda le sottoregioni formanti il profilo recettore delle cellule corticali, questo non è più in genere circolare, ma tende a diventare ellittico (figura 1.11b); di conseguenza lo stimolo migliore per attivare la cellula non è più un puntino luminoso ma una linea, una barra o un margine chiaro-scuro (edge).

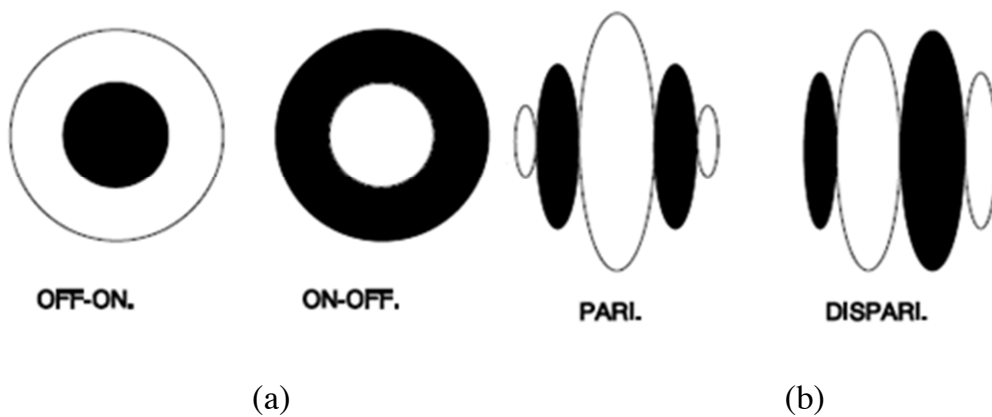


Figura 1.11:

(a) Rappresentazione schematica dei profili recettori di neuroni della retina o del LGN. (b) Rappresentazione schematica dei profili recettori di neuroni della corteccia visiva primaria.

Inoltre emerge nella corteccia anche la necessità che lo stimolo luminoso abbia una determinata orientazione spaziale [3]. Ciò vuol dire ad esempio che uno stimolo luminoso, per contribuire in maniera efficace all'aumento di attività cellulare, deve avere una determinata orientazione rispetto alle coordinate principali del campo visivo. Ogni cellula ha una sua orientazione preferenziale e la sua capacità di rispondere diminuisce a mano a mano che l'orientazione dello stimolo si discosta da quello ottimale. Così, se una linea posta ad esempio a $\pi/4$ rispetto alla verticale rappresenta uno stimolo

efficace quando cade nel profilo recettore di una determinata cellula corticale, nel momento in cui la sua inclinazione viene cambiata a $\pi/6$ questa la stessa cellula corticale non sarà influenzata con la stessa forza. La preferenza per stimoli allungati e con un certo orientazione è una caratteristica comune delle cellule di V1.

Hubel e Wiesel (premi Nobel per le scoperte sul funzionamento della corteccia visiva primaria nel 1981) hanno distinto le cellule della corteccia visiva primaria in due categorie sulla base di altre proprietà di risposta, cellule semplici e complesse. [13] [14].

Secondo l'analisi di Hubel e Wiesel i profili recettori delle cellule semplici sono suddivisi in due o tre zone parallele "ON" e "OFF" ben definite. Tuttavia il numero di sottoregioni parallele all'interno del profilo recettore, che in genere è di due o tre, può essere anche superiore [8][2]. I profili che si ottengono dalle misure sperimentali sono ben approssimabili con delle funzioni di Gabor (sezione 1.2.1), che, nel macaco, possono avere simmetria pari o dispari al variare della fase (figura 1.11b). Le cellule con profili recettori pari e dispari sono dette cellule semplici [26]. Esse sono sensibili alla frequenza spaziale dello stimolo (quindi alla grandezza dei lobi e alla larghezza dello stimolo ottimo), alla fase e all'orientazione [8].

Come le cellule semplici, le cellule complesse sono sensibili alla frequenza spaziale dello stimolo e alla sua orientazione ma non alla fase. Si ritiene che la loro risposta sia l'esito dell'integrazione non lineare della risposta di due cellule semplici che hanno medesime frequenza spaziale e orientazione preferite di quelle della cellula complessa ma differiscono nella fase di $\pi/2$ (ossia una è pari e l'altra è dispari; sezione 1.2.1) [1]. Come agiscono i profili recettori? Ossia quale informazione sullo stimolo fornisce l'attivazione di un'unità con un determinato profilo recettore avente particolari frequenza, ampiezza ed orientazione preferiti? In figura 1.12 è schematizzata la posizione relativa tra un profilo recettore dispari ed uno pari ed un contorno. Si noti che la risposta dell'unità legata al profilo recettore dispari è massima in valore assoluto (quindi massima o minima)

quando l'asse di simmetria del profilo recettore, e, quindi, il punto spaziale che rappresenta il centro del profilo recettore, sono su un bordo, cioè su una linea che rappresenta un alto contrasto; si noti inoltre che la risposta è minima quando il profilo recettore copre regioni di luminanza uniforme (lo sfondo o la figura). Per un profilo recettore pari, invece, la risposta è massima quando l'asse di simmetria del profilo recettore è al centro di una barretta orientata mentre è minima quando si trova lungo linea che presenta alto contrasto. I profili recettori dispari sono quindi rilevatori di bordi. I profili recettori pari per la loro conformazione, invece, rispondono in maniera preferenziale a barrette orientate.

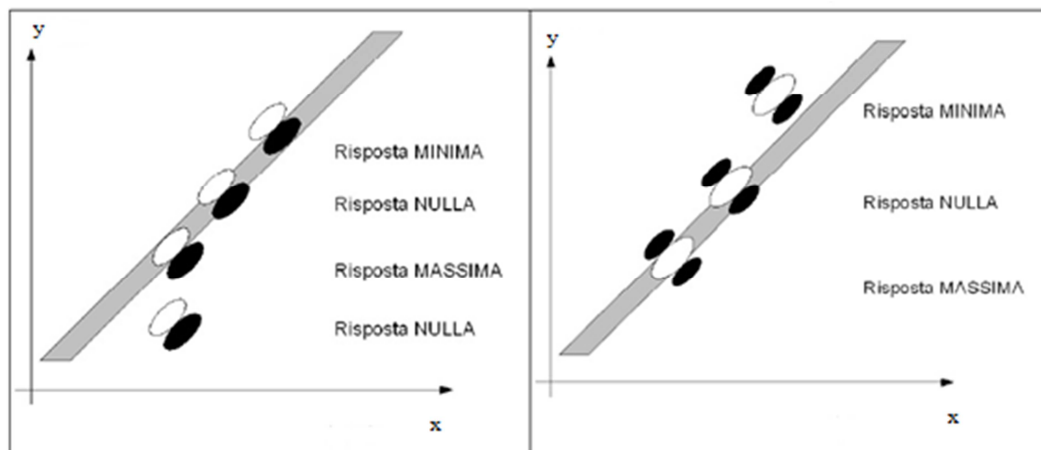


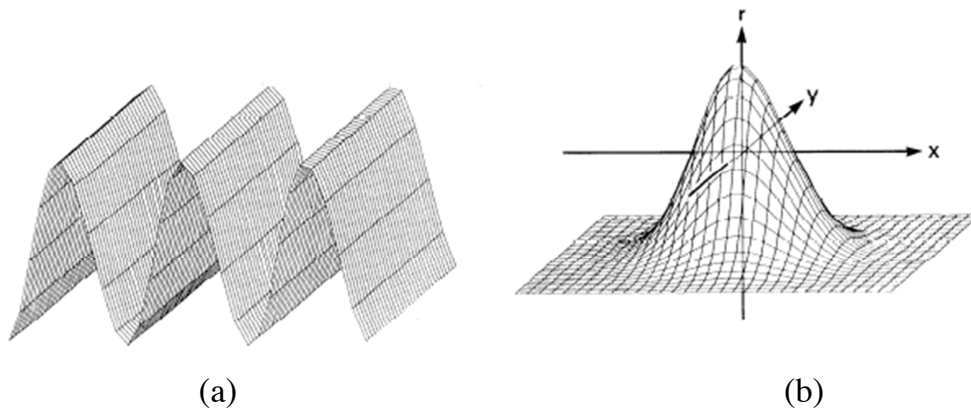
Figura 1.12:

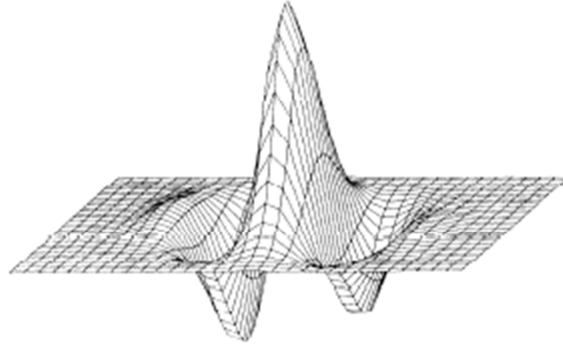
Esempio di risposta cellulare in base alla posizione relativa di un profilo recettore (dispari a sinistra, pari a destra) ed un contorno.

Le cellule complesse hanno profili recettori in genere più grandi delle semplici, ma non ci sono zone “ON” e “OFF” chiaramente separate. Inoltre la maggior parte delle complesse risponde in modo più marcato a stimoli in movimento ed il movimento deve avvenire in direzione ortogonale all’orientazione (selettività alla direzione di spostamento). Quando lo stimolo luminoso invade tali zone si ha una riduzione od abolizione della scarica.

1.2.1 I filtri di Gabor

Come illustrato in precedenza, le cellule semplici della corteccia visiva rispondono vigorosamente a linee o bordi di dato orientazione oltre a stimoli con particolare frequenza spaziale. Per considerare entrambe queste caratteristiche occorre perciò applicare la teoria dei filtri, assumendo che il sistema presenti caratteristiche lineari, come visto nella sezione precedente. Nel 1946 Gabor ha definito, nel suo fondamentale lavoro “Theory of communication”, il principio di indeterminazione applicato alla teoria dei segnali, secondo il quale un’elevata risoluzione spaziale determina una perdita di risoluzione in frequenza di un filtro. Gabor ha introdotto la famiglia di funzioni, a cui è stato associato il nome, che minimizza l’indeterminazione tra la risoluzione nel tempo e in frequenza, che può essere utilizzata per realizzare filtri bidimensionali [11]. Dall’analogia tra la struttura dei profili recettori e le funzioni di Gabor, determinata sperimentalmente [17], è possibile affermare che questi filtri si adattano a descrivere i profili recettori di cellule semplici. Questo modello ha il grande vantaggio di possedere alcune tra le maggiori proprietà dei profili recettori di neuroni appartenenti a V1: struttura del profilo recettore nello spazio, selettività in orientazione e in frequenza spaziale. Un filtro di Gabor è un filtro lineare, ottenuto dal prodotto di una gaussiana e di un’onda piana sinusoidale orientata.





(c)

Figura 1.13:

Struttura di un filtro di Gabor bidimensionale nello spazio. Questo è rappresentato in (c) e può essere descritto come il prodotto di un'onda piana sinusoidale (a) e una funzione Gaussiana (b).

Un filtro di Gabor bidimensionale nel dominio spaziale (x,y) ha la seguente espressione:

(1.3)

$$g_{\alpha,\beta,p_0,p_0}(x,y) = e^{-\pi[\alpha^2(x-x_0)^2+\beta^2(y-y_0)^2]} e^{-i[2\pi(u_0(x-x_0)+v_0(y-y_0))+\varphi]}$$

da cui:

$$(1.4) \quad \begin{aligned} \operatorname{Re}\{g(x,y)\} &= e^{-\pi[\alpha^2(x-x_0)^2+\beta^2(y-y_0)^2]} \cos(2\pi(u_0(x-x_0)+v_0(y-y_0))+\varphi) \\ \operatorname{Im}\{g(x,y)\} &= e^{-\pi[\alpha^2(x-x_0)^2+\beta^2(y-y_0)^2]} \sin(2\pi(u_0(x-x_0)+v_0(y-y_0))+\varphi) \end{aligned}$$

dove la parte reale di una funzione di Gabor descrive i profili recettori di cellule semplici con simmetria pari, mentre la parte immaginaria i profili recettori di cellule dispari [7], e queste coppie di cellule sono in quadratura di fase. La funzione di Gabor (1.3) è quindi costituita da una gaussiana

bidimensionale moltiplicata per un'onda piana, dove la gaussiana è ottenibile come il prodotto tra due gaussiane monodimensionali, una lungo l'asse x , di larghezza α e centrata in x_0 , l'altra lungo l'asse y , di larghezza β e centrata in y_0 .

Uno dei problemi nell'utilizzare cellule semplici come filtri è dato dalla sensibilità di questi alla variazione di fase: come descritto in precedenza, i loro profili recettori presentano regioni aventi opposta polarità e la loro risposta a uno stimolo dipende da come questo si allinea ai profili recettori. Per ottenere informazioni indipendenti dalla fase dello stimolo, occorre considerare le cellule complesse che prendono in ingresso input provenienti da una coppia di cellule semplici con la stessa orientazione preferenziale del profilo recettore ma in quadratura di fase tra loro (espressioni riportate in 1.4) [1]. Sommando i quadrati dei due output provenienti da due cellule semplici in quadratura di fase si ottiene l'energia di moto: questa energia sarà sempre positiva e tanto più elevata quanto più l'orientazione dello stimolo sarà solidale con quella delle coppie di cellule semplici. Considerando infatti una coppia di cellule semplici, una con simmetria pari e l'altra dispari, è possibile affermare che l'output di queste sarà sensibile alla variazione di fase; sommando però il quadrato di ciascuno dei due output risulta possibile estrarre l'energia di moto. Una funzione di Gabor per cellule complesse è quindi in grado di approssimare e modellare i profili recettori di cellule corticali, evidenziando che questi neuroni sono in grado di rilevare la frequenza spaziale e l'orientazione di uno stimolo in ingresso minimizzando l'indeterminazione anche sull'orientazione Θ [7].

Come verrà illustrato nel capitolo 2, la modellazione dei profili recettori con una funzione di Gabor è di fondamentale importanza per la definizione di un modello differenziale.

1.3 Organizzazione della corteccia visiva

Nel corso degli ultimi decenni, numerosi sono stati gli studi sperimentali volti a spiegare l'organizzazione funzionale della corteccia visiva. Gli esperimenti consistono nel misurare l'attività neurale con uno o più microelettrodi, mentre questi penetrano la corteccia in direzione perpendicolare o quasi tangenziale ad essa. Hubel e Wiesel [13] [14], per primi, riscontrarono un'organizzazione ordinata nella corteccia visiva, secondo quanto fu da essi osservato, penetrando la corteccia in direzione perpendicolare, si incontrano cellule semplici e complesse accomunate dalla stessa posizione retinica del profilo recettore e dalla stessa propensione per un determinato orientazione; a seguito di una penetrazione obliqua (quasi orizzontale) si registrano, invece, cambiamenti sequenziali dell'orientazione preferenziale.

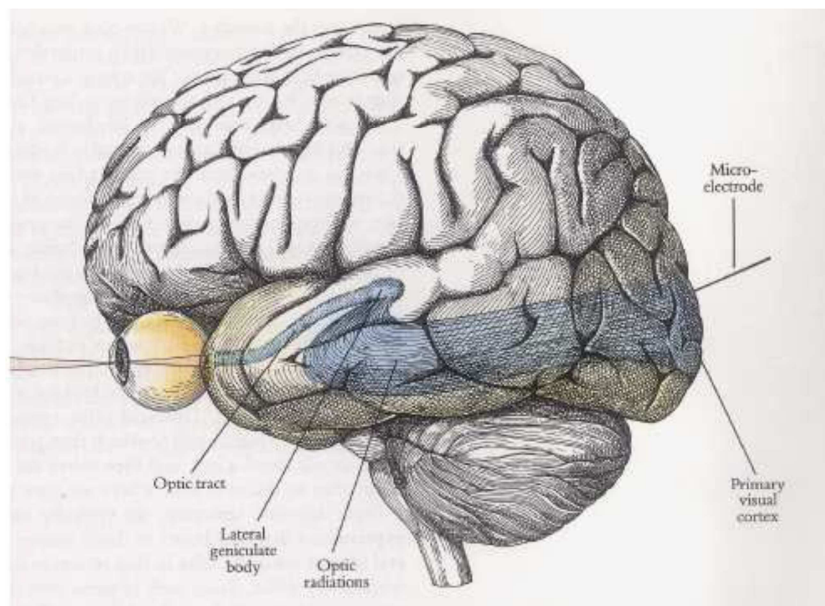


Figura 1.14:

Schema del percorso affrontato dal segnale contenente l'informazione visiva.

Inoltre, sebbene, per piccoli spostamenti sulla corteccia, le posizioni dei campi recettori siano soggette a fluttuazioni casuali, esiste una stretta relazione tra la direzione dello spostamento sulla corteccia e quella nel campo visivo che consente la conservazione della topologia dell'immagine. L'unione di questi risultati ha condotto Hubel e Wiesel ad interpretare la corteccia visiva come una organizzazione periodica di unità funzionali, dette colonne, contenenti neuroni selettivi ad un particolare orientazione. L'insieme di colonne verticali corrispondenti ad una sequenza completa di orientazioni (un periodo), secondo la terminologia adottata da Hubel e Wiesel, prende il nome di ipercolonna. L'ipercolonna rappresenta quindi uno dei blocchi costitutivi fondamentali della corteccia.

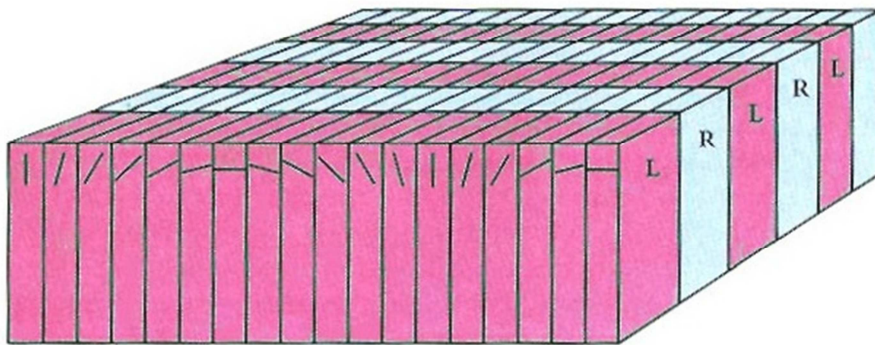


Figura 1.15:

Organizzazione della corteccia visiva primaria in colonne e ipercolonne.

La corteccia visiva primaria, così come altre aree associative, presenta quindi un'organizzazione in moduli (o colonne) che si ripetono periodicamente.

Queste colonne sono organizzate nello spazio secondo una variazione di orientazione minima e continua, interrotta solo da strutture chiamate "pinwheels", in cui ogni colore rappresenta l'orientazione dello stimolo che attiva maggiormente ogni regione della corteccia (la mappa di orientazione preferenziale e struttura a "pinwheels" verranno illustrate nel capitolo 3).

Nell'area striata ogni modulo è caratterizzato da neuroni sensibili alle frequenza spaziale, al movimento e a linee o bordi di particolare orientazione all'interno di una specifica porzione del campo visivo. All'interno di ogni modulo le caratteristiche di risposta delle cellule hanno una disposizione sistematica. La struttura principale è rappresentata dalle colonne di dominanza oculare, all'interno della quale sono rappresentati, con pesi variabili, progressivamente e ordinatamente neuroni che ricevono input solo da uno dei due occhi o da entrambi. Interposta a questa struttura colonnare è presente una struttura, detta dei blobs (selettivi per il colore), centrata all'interno della colonna stessa.

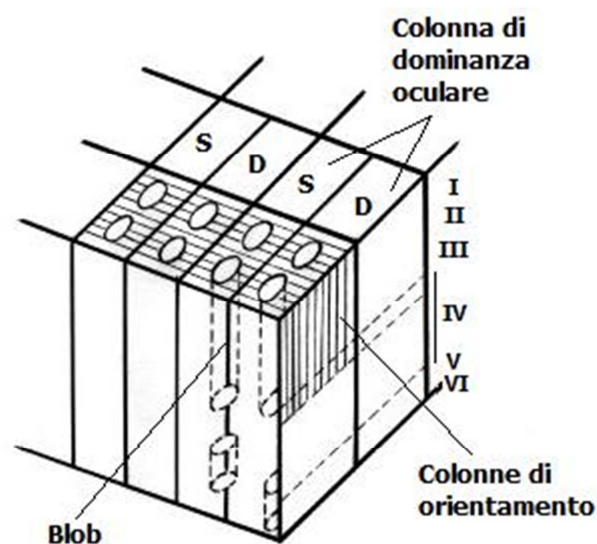


Figura 1.16: Organizzazione di V1.

Le cellule appartenenti al medesimo modulo corticale – ve ne sono sia di semplici sia di complesse – si distinguono quindi per l'orientazione e la frequenza. È interessante notare che all'interno di una ipercolonna, le unità sono funzionalmente organizzate in modo che cellule con profili recettori di simile orientazione siano vicine tra di loro [4].

La colonna accede ad una regione del campo visivo corrispondente ad un aggregato dei profili recettori in essa contenuto. Più precisamente, è

necessario uno spostamento sulla corteccia di circa 2.4 mm, affinché due colonne accedano a regioni completamente distinte dell'immagine.

La struttura ipercolonnare organizza perciò le cellule corticali in colonne corrispondenti al parametro di orientazione: ad ogni posizione sulla retina è associato un insieme di cellule, l'ipercolonna, costituito da tutte le cellule sensibili a tutte le possibili orientazioni e centrate nel punto fissato [15].

Nel prossimo capitolo verrà presentato il concetto di campo di associazione per stimoli orientati, definito attraverso esperimenti psicofisiologici da Field, Hayes e Hess. Attraverso i campi di associazione sarà possibile definire quali caratteristiche due segmenti devono avere per essere associati allo stesso gruppo percettivo.

Verrà inoltre introdotto il modello di campi di associazione in un dominio che estende il piano immagine \mathbb{R}^2 per poter modellare la connettività cellulare. L'elaborazione degli stimoli infatti non viene fatta soltanto sul piano bidimensionale dell'immagine, il cervello tende ad aumentare la dimensionalità dello stimolo visivo, aggiungendo come coordinata l'orientazione (si passa dal dominio \mathbb{R}^2 dell'immagine bidimensionale a $\mathbb{R}^2 \times S^1$).

Capitolo 2

La percezione visiva:

dalle teorie della Gestalt al modello differenziale Citti-Sarti

La percezione fa parte dei processi cognitivi umani (come la memoria, l'intelligenza, l'apprendimento) e rappresenta un'elaborazione da parte del nostro organismo degli stimoli che provengono dall'ambiente esterno.

Le sensazioni e le percezioni non esistono quindi come attività isolate, ma fanno parte di un processo più ampio di esperienza e conoscenza che prende il nome di attività cognitiva. Per sistema percettivo si intende l'attività degli organi di senso e l'organizzazione degli impulsi provenienti dai sensi da parte del sistema nervoso. Il tema che caratterizza questo capitolo è quello della percezione visiva. All'inizio del secolo scorso la corrente ideologica della Gestalt (dal tedesco: forma, schema, rappresentazione) inizia a definire delle regole secondo le quali stimoli visivi distinti tra loro possono essere percepiti come un'unica unità percettiva. Per la psicologia della Gestalt non è giusto dividere l'esperienza umana nelle sue componenti elementari e occorre invece considerare l'intero come fenomeno sovraordinato rispetto alla somma dei suoi componenti: "L'insieme è più della somma delle sue parti".

L'idea di fondo è che i fenomeni complessi possono essere compresi solo utilizzando l'idea di struttura, piuttosto che quella di elemento. Per comprendere il mondo circostante si tende a identificarvi forme secondo schemi che sembrano adatti e scelti per imitazione, apprendimento e condivisione; attraverso simili processi si organizzano la percezione, il pensiero e la sensazione [27].

La nascita della psicologia della Gestalt risale esattamente al 1912, quando Max Wertheimer [30] scrive un articolo in cui identifica un processo percettivo unitario grazie al quale i singoli stimoli vengono integrati, nel

soggetto, in una forma dotata di continuità. Il percepire, prima considerato un processo passivo, viene ad essere pensato come qualcosa di gran lunga più attivo: come un'attività subordinata a certi principi organizzativi generali. La percezione dunque non dipende dagli elementi ma dalla strutturazione di questi in un "insieme organizzato", in una forma [27].

In riferimento alla percezione visiva, i principi della Gestalt hanno perciò il fine di spiegare quali parti dello stimolo sono associate formando unità percettive salienti (bordi, figure, sfondi, qualità cromatiche e acromatiche).

Le regole principali della percezione visiva individuate dalla Gestalt sono:

- la prossimità: gli elementi sono raggruppati in funzione della distanza relativa degli elementi che compongono la figura;
- la somiglianza: gli elementi vengono uniti in forme con tanta maggior coesione quanto maggiore è la loro somiglianza;
- il destino comune: dato un insieme di elementi in movimento, vengono raggruppati quelli con spostamento coerente. In una configurazione tendono a unificarsi le linee con la stessa direzione od orientazione o movimento, secondo l'andamento più coerente, a difesa delle forme più semplici e più equilibrate;
- la buona continuazione: tutti gli elementi sono percepiti come appartenenti ad un insieme continuo e coerente, si tendono perciò a percepire come unità quegli elementi che minimizzano i cambiamenti di direzione;
- la chiusura: le linee che formano delle figure chiuse tendono ad essere viste come unità formali;
- l'esperienza passata: elementi che per la nostra esperienza passata sono abitualmente associati tra di loro tendono ad essere uniti in forme.

Questi concetti di carattere prettamente psicologico più che fisiologico o scientifico si basano su come la percezione del mondo esterno possa influenzare il comportamento umano [27].

Ha suscitato notevole interesse anche il tema riguardante la percezione dei contorni. Ci sono due tipi di completamento che il cervello umano mette in atto: il completamento modale, ovvero quello che avviene quando le parti interpolate dell'immagine sono percepite con la piena modalità della visione e sono fenomenologicamente indistinguibili dallo stimolo reale (come la creazione di contorni illusori) e il completamento amodale, quel tipo di processo attraverso cui il cervello ricostruisce i contorni e la configurazione di un oggetto parzialmente occluso.

Una dimostrazione efficace di entrambi i tipi di completamento è il triangolo di Kanizsa (figura 2.1). Descritto per la prima volta nel 1955 da Gaetano Kanizsa, uno dei principali esponenti della scuola gestaltista italiana, il triangolo rappresenta uno dei fenomeni percettivi più indagati.

Osservando la figura il nostro sistema visivo è in grado di percepire non soltanto i contorni effettivamente presenti, ma anche altri contorni, detti contorni soggettivi o illusori. Infatti noi percepiamo molto chiaramente la presenza di un triangolo bianco centrale, che occlude tre dischi neri e un triangolo a tratto, nonostante i bordi di questa figura non siano completamente presenti nell'immagine. Inoltre la regione bianca ci appare come una superficie in rilievo, più chiara dello sfondo, e si chiama superficie soggettiva perchè ricostruita dal soggetto che guarda.

Questo esempio comprende dunque sia un esempio di completamento modale, la comparsa del triangolo centrale, sia ben quattro casi di completamento amodale: quello dei tre dischi, e quello del triangolo incompleto che si colloca anch'esso dietro al triangolo centrale occludente [18].

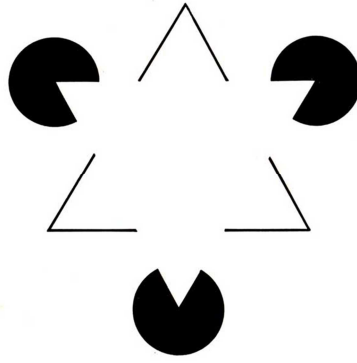


Figura 2.1: Triangolo di Kanizsa.

Altri studi sulla percezione visiva hanno richiamato in causa le teorie della Gestalt, confermate, almeno a grandi linee, dalle analisi effettuate sul funzionamento dei neuroni della corteccia visiva primaria. In particolare, il classico lavoro di Field, Hayes e Hess del 1993 [9], ha quantificato un'analisi che per la Gestalt era qualitativa, in particolare le regole di prossimità e di buona continuazione, attraverso la definizione dei campi di associazione per stimoli orientati.

Numerosi modelli sono stati proposti, durante gli anni, per cercare di interpretare la modalità con la quale le cellule corticali sono connesse tra loro.

Verranno ora presentati la teoria dei campi di associazione di Field, Hayes ed Hess ed il modello differenziale Citti-Sarti, che giustifica il completamento percettivo sulla base della funzionalità della corteccia.

2.1 I campi di associazione

Come descritto in precedenza, le leggi della Gestalt costituiscono delle regole per raggruppare stimoli visivi distinti tra loro in un'unica unità percettiva. In particolare, la legge della buona continuazione è stata usata per rappresentare una varietà di fenomeni che dimostrano l'importanza della continuità di parametri visivi nella percezione umana.

Nel 1993 Field, Hayes e Hess hanno definito, attraverso esperimenti psicofisiologici, dei *campi di associazione* per stimoli orientati, cercando di quantificare ciò che gli psicologi avevano descritto qualitativamente, con lo scopo di individuare le regole che governano la percezione della continuità. Attraverso cinque esperimenti hanno definito le caratteristiche che gli stimoli orientati devono avere per poter essere associati allo stesso gruppo percettivo; questi hanno permesso di capire come l'allineamento di elementi vicini, ognuno rappresentato da una funzione di Gabor, è correlato alla percezione della continuità. Gli esperimenti sono tutti riconducibili ad uno stesso schema di seguito riassunto. Per una durata di circa un secondo, sono state presentate ai soggetti (i due autori) due immagini, una con gli elementi dello sfondo, che presentano un'orientazione casuale, l'altra contenente sia gli elementi dello sfondo che il percorso di interesse: 12 elementi disposti in modo che i rispettivi assi principali siano collineari ad una linea curva continua immaginaria (non fisicamente presente ma che può essere percepita per completamento) [9].

L'unità percettiva è stata creata usando le regole descritte in figura 2.2 e viene visualizzata in figura 2.3.

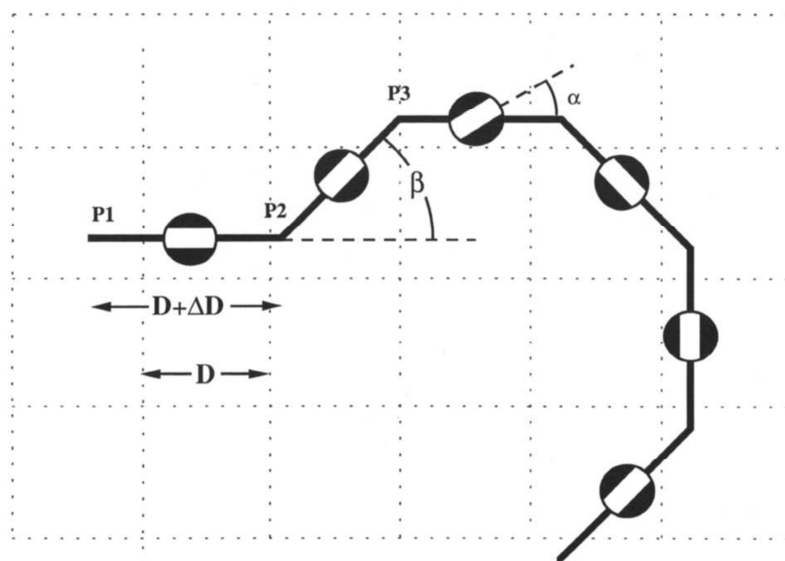


Figura 2.2:

Griglia che mostra la creazione del percorso di interesse. Ogni stimolo (costituito da 256 elementi) viene diviso in una griglia di 16x16 quadrati, ognuno contenente un elemento: 12 appartenenti al percorso, 244 allo sfondo. Gli elementi dello stimolo sono stati posti entro i quadrati lungo un percorso predeterminato. In figura, D rappresenta la lunghezza del segmento entro il quale un elemento deve cadere, l'angolo β è la differenza di orientazione tra elementi successivi (solitamente compreso tra ± 5 gradi), infine l'angolo α è l'angolo di orientazione della sinusoide rispetto al percorso (nella maggior parte delle prove l'orientazione degli elementi è stata posta uguale all'orientazione dello stesso, quindi $\alpha=0$).

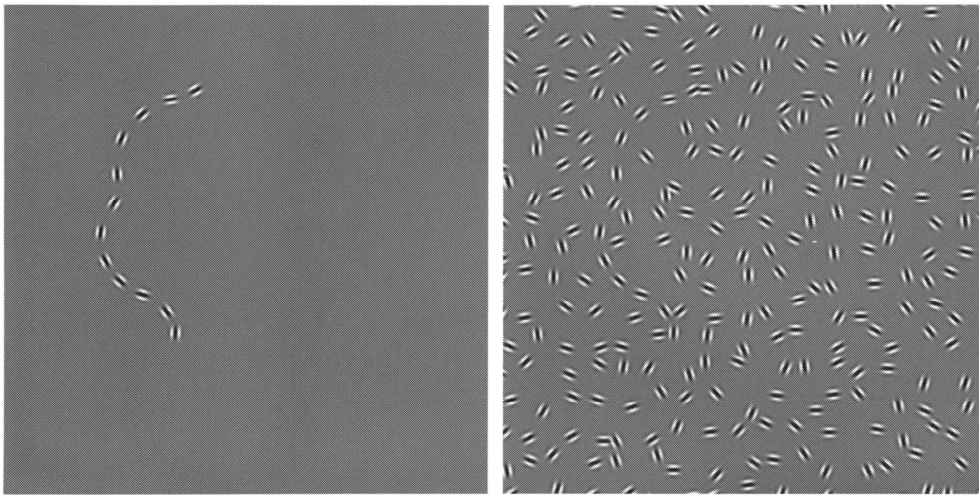


Figura 2.3:

Esempio di stimolo usato negli esperimenti. A sinistra è riportato il percorso che il soggetto è stato chiamato a riconoscere quando esso è posto all'interno di un insieme di segmenti orientati in maniera casuale (figura di destra).

Il compito del soggetto è stato quello di identificare lo stimolo in cui era presente la serie di elementi disposti in modo da formare il percorso di

interesse. Le variabili in gioco erano: l'orientazione relativa e l'allineamento tra elementi successivi, la relazione angolare tra gli assi di simmetria degli elementi e il percorso, la distanza tra gli elementi e la loro fase. I risultati principali che sono stati conseguiti attraverso questi esperimenti sono di seguito riportati:

- maggiore è l'angolo di orientazione tra elementi successivi più è difficile riconoscere il percorso: il completamento della linea target avviene per elementi che hanno orientazioni che differiscono fino a $\beta = 60^\circ$;
- nel caso di differenza di orientazione tra elementi successivi (allineamento "side-to-side", a differenza del "end-to-end" precedente) risulta più difficile identificare il percorso: l'allineamento gioca quindi un ruolo fondamentale nella percezione della continuità (figura 2.4);
- variando l'orientazione degli elementi rispetto all'orientazione del percorso di un angolo α , questo risulta più difficile da identificare;
- una distanza elevata tra gli elementi riduce l'abilità dell'osservatore di identificare il percorso, nonostante ciò questa può valere anche fino a 7 volte la loro dimensione: l'integrazione risulta possibile anche a grandi distanze;
- la fase degli elementi è irrilevante rispetto alla loro distanza ed alla loro orientazione.

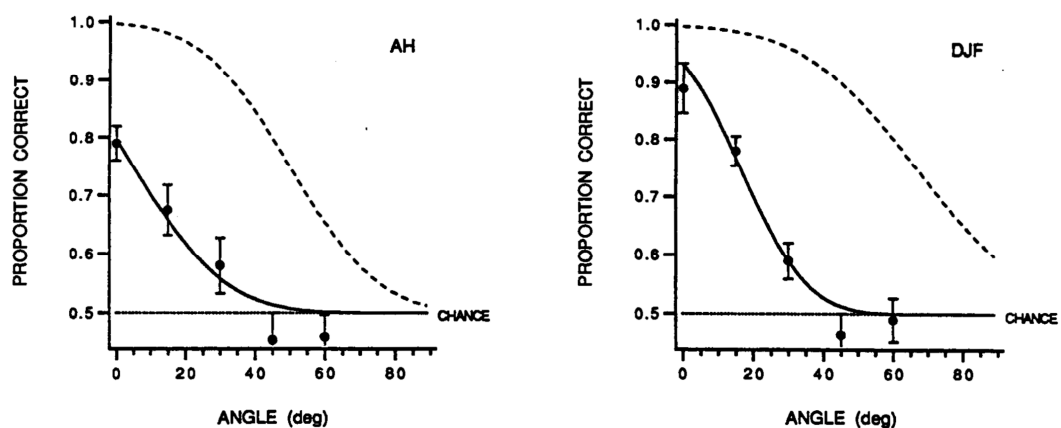


Figura 2.4:

Risultati della seconda prova per i due soggetti. I cerchi neri rappresentano i valori ottenuti, mentre la linea continua una funzione che li interpola e in questo caso rappresenta il grado di individuazione della traiettoria quando gli elementi sono ortogonali al percorso. La linea tratteggiata, ricavata dal primo esperimento, viene utilizzata come confronto per i risultati ottenuti nelle prove e rappresenta il grado di individuazione della traiettoria quando gli elementi sono collineari al percorso. L'associazione risulta quindi più debole quando elementi successivi non sono allineati tra loro in maniera "end-to-end", ma "side-to-side".

Complessivamente questi risultati hanno portato a postulare l'esistenza di un campo di associazione, presente in ogni punto dello spazio visivo bidimensionale, che permette di legare elementi orientati tra loro distanti, secondo tre variabili principali: la curvatura della linea di associazione, la sua lunghezza e la relazione tra l'asse principale degli elementi da associare e la direzione della curva nel punto in cui incontra gli elementi stessi. La figura 2.5 mostra una schematizzazione dei campi di associazione: dato un elemento, esiste una regione attorno ad esso dove gli altri elementi vengono associati e separati dallo sfondo. Le prove sperimentali hanno mostrato che i campi di associazione dipendono dall'orientazione dell'elemento originante e i restanti stimoli visivi vengono associati a seconda della loro orientazione e della loro posizione relativa a quest'ultimo [9].

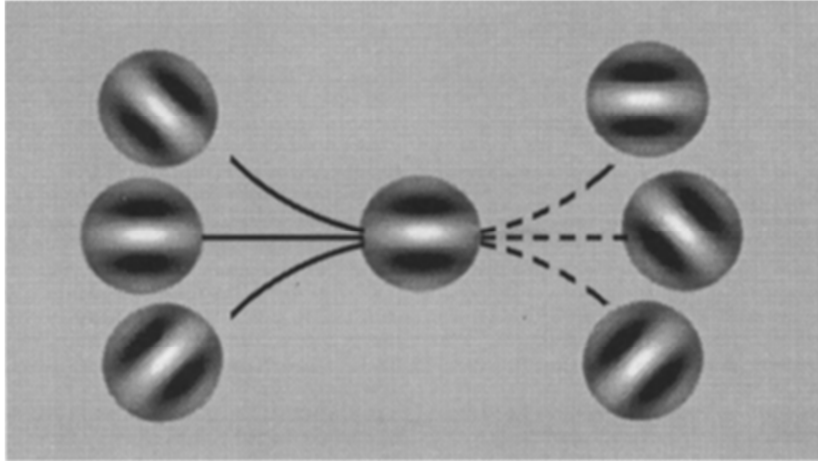


Figura 2.5:

Campi di associazione. Le curve in figura sono una schematizzazione delle regole di allineamento. Esiste una forte correlazione tra posizione e orientazione: mentre gli elementi con allineamento come quello a sinistra della figura sono “associati”, gli altri non lo sono, nonostante le differenze di orientazione siano le stesse.

Questo studio ha permesso di determinare alcuni dei parametri che influiscono sull’abilità dell’uomo di separare uno stimolo dal background e le regole dei campi di associazione che definiscono quali caratteristiche due segmenti devono avere per poter essere associati allo stesso gruppo percettivo.

Un modello con il quale è possibile modellare la struttura di campi di associazione e che giustifica il completamento percettivo sulla base della funzionalità della corteccia, è stato sviluppato da G. Citti e A. Sarti [5]. Tramite questo è possibile ricavare una famiglia di curve integrali utilizzabili per modellare i campi di associazione.

2.2 Modello differenziale Citti-Sarti per il completamento percettivo

Per capire come il sistema visivo interpola le informazioni esistenti e forma le unità percepite sono stati sviluppati modelli matematici di completamento percettivo. Questi modelli prendono in considerazione molte proprietà fenomenologiche, come descritto dagli psicologi della Gestalt. Dal punto di vista neurofisiologico, il completamento viene compiuto dal primo strato della corteccia visiva, attraverso due meccanismi a cascata: l'uno estrae l'informazione (contorni reali, gradienti dell'immagine) tramite un filtraggio feed-forward, l'altro la completa tramite un circuito ricorsivo. Il primo meccanismo viene compiuto dalle cellule semplici della corteccia primaria che estraggono informazioni sul modulo e sull'orientazione della luminosità del gradiente dello stimolo visivo. Il secondo propaga l'informazione estratta in funzione dell'orientazione, in termini di connessioni orizzontali. Questa connettività a lungo raggio assicura la connessione tra cellule che hanno la stessa orientazione ma appartenenti a diverse ipercolonne [5]. Come verrà illustrato nel prossimo capitolo, Bosking et al hanno proposto la possibilità che le connettività orizzontali a lungo raggio servano a implementare i campi di associazione [4], dove la forza del collegamento sinaptico dipende dalla posizione e dall'orientazione delle cellule connesse.

La propagazione dell'informazione può essere descritta dai campi di associazione, che ne indicano la direzione privilegiata di diffusione. In questa sezione viene presentato il modello di campi di associazione nello spazio $\mathbb{R}^2 \times S^1$ di Citti-Sarti, che introduce una giustificazione del completamento percettivo sulla base della funzionalità della corteccia visiva primaria (V1).

Come mostrato in precedenza, le cellule semplici possono essere considerate unità sensibili al gradiente luminoso, oltre che alla frequenza dello stimolo, alla fase e all'orientazione [8]. I profili recettori di queste

cellule corticali vengono modellati da filtri bidimensionali di Gabor, la cui espressione è riportata in (1.3). Riprendendo il parte la notazione del capitolo precedente, un filtro di Gabor con orientazione θ ha la seguente espressione:

$$(2.1) \quad G(x, y, \theta) = \frac{1}{2\pi s^2} * e^{-\frac{(\tilde{x}^2 + \tilde{y}^2)}{s^2} + i\tilde{y}/s}$$

dove:

$$\tilde{x} = x\cos(\theta) + y\sin(\theta) \text{ e } \tilde{y} = -x\sin(\theta) + y\cos(\theta).$$

Queste condizioni descrivono una rotazione degli assi di un angolo θ , così che i filtri di Gabor sono ottenuti da una funzione fissata, tramite una rotazione del dominio [5].

La parte immaginaria del filtro, cioè la sua parte dispari, può essere approssimata a:

$$(2.2) \quad 2 \frac{\sin(\tilde{y})}{s^2} e^{-(\tilde{x}+\tilde{y})/s^2} \simeq \frac{2\tilde{y}}{s^2} e^{-(\tilde{x}+\tilde{y})/s^2} = -\partial_{\tilde{y}} e^{-(\tilde{x}+\tilde{y})/s^2} .$$

Il funzionamento di una cellula semplice modellata attraverso la parte immaginaria di un filtro di Gabor, approssimato da una derivata direzionale, può essere interpretato come un operatore differenziale che agisce sull'immagine.

Una derivata in direzione \tilde{y} può essere espressa nelle variabili originali (x, y, θ) come una derivata direzionale, lungo il vettore $(-\sin \theta, \cos \theta)$:

$$(2.3) \quad X_3 = -\sin\theta\partial_x + \cos\theta\partial_y$$

Questa derivata, applicata all'immagine definita da una funzione I , esprime la proiezione del gradiente nella direzione $(-\sin \theta, \cos \theta)$. Con questa

notazione il filtraggio genera, attraverso la convoluzione con I , una funzione:

$$(2.4) \quad O(x, y, \theta) = -X_3 e^{-(\tilde{x}+\tilde{y})/s^2} * I = -X_3(\theta)I_s$$

dove

$$(2.5) \quad I_s = I * e^{-(\tilde{x}+\tilde{y})/s^2}$$

La funzione $O(x, y, \theta)$ dipende in ogni punto sul piano Immagine (x, y) dall'orientazione locale θ e decade in maniera esponenziale dai punti di valore massimo, comportamento dovuto alla presenza nell'espressione della funzione Gaussiana. Questa funzione, per θ fissato, seleziona punti vicini dove la componente del gradiente in direzione $(-\sin \theta, \cos \theta)$ è sufficientemente elevata, come mostrato in figura 2.6:

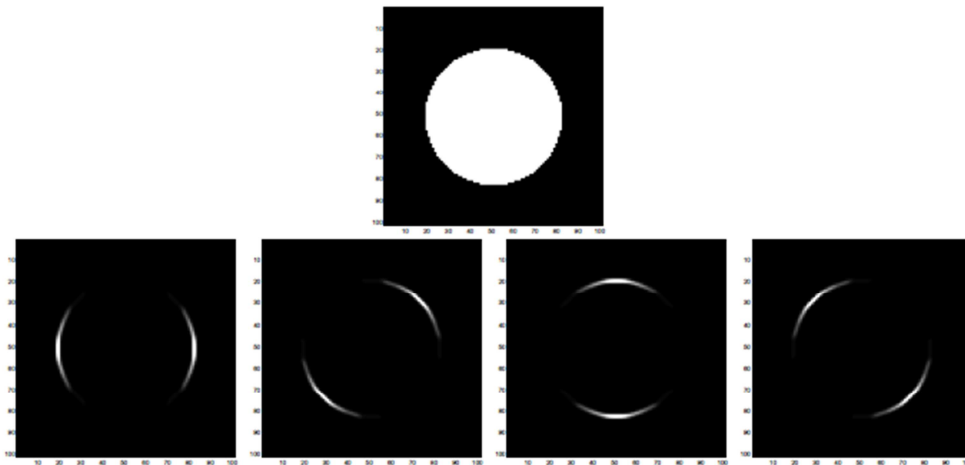


Figura 2.6:

In alto è riportata l'immagine originale, un disco bianco, e in basso una serie di sue convoluzioni con filtri di Gabor aventi diverse orientazioni, ovvero fette della funzione $O(x, y, \theta)$.

È possibile identificare il dominio su cui vive la funzione O con le fibre di orientazione che partono da ogni punto sul piano Immagine: ogni fibra rappresenta un'ipercolonna dove le cellule corticali, rappresentate nelle loro diverse features spazio-orientazione da tutti i punti del dominio $\mathbb{R}^2 \times S^1$ e sensibili alla posizione sul piano Immagine, sono disposte in funzione dell'orientazione locale, come mostrato in figura 2.7.

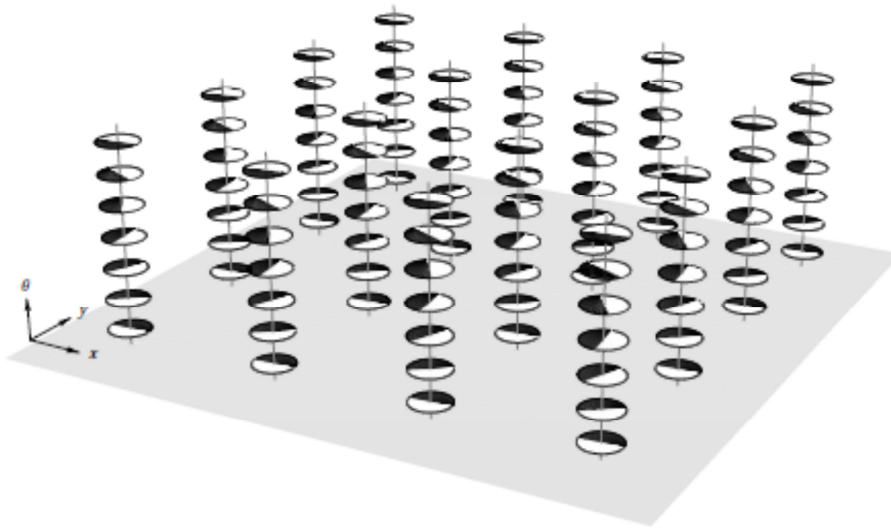


Figura 2.7:

Organizzazione di V1 in ipercolonne. Formalmente sopra ad ogni punto (x,y) del piano Immagine è presente un'intera ipercolonna di cellule corticali, ciascuna individuata da un'orientazione locale θ .

Questo meccanismo di convoluzione è però insufficiente per spiegare l'alta selettività all'orientazione delle cellule semplici e complesse in V1. Meccanismi fisiologici [5], come ad esempio il feedback neurale, operano per ogni punto la selezione dell'orientazione del massimo output dell'ipercolonna in risposta allo stimolo visivo, eliminando tutti gli altri output e individuando così l'orientazione di massima risposta delle cellule in V1. Definito $X_3(\theta)I_s$ la proiezione del gradiente lungo la direzione del

vettore $(-\sin \theta, \cos \theta)$, il massimo sarà raggiunto al valore $\bar{\theta}$, che rappresenta la direzione del gradiente. Si ottiene così:

$$(2.6) \quad |X_3(\bar{\theta})| = \max_{\theta} |X_3(\theta)|.$$

In questo processo ogni punto del dominio 2D dell'immagine è 'liftato' al punto $(x, y, \bar{\theta})$. Aggiungendo come coordinata l'orientazione si passa quindi dal dominio \mathbb{R}^2 dell'immagine bidimensionale a $\mathbb{R}^2 \times S^1$.

L'insieme dei valori 'liftati' corrisponde al massimo dell'attività dell'output delle cellule in V1.

Con X_3 quindi si indentifica una sezione dello spazio tangente a $\mathbb{R}^2 \times S^1$, il quale ha dimensione 3 in ogni punto dello spazio euclideo. È possibile dunque definire dei piani orizzontali bidimensionali, sottoinsieme dello spazio tangente $\mathbb{R}^2 \times S^1$, ortogonale a X_3 , generati dai seguenti campi vettoriali:

$$(2.7) \quad X_1(\theta) = (\cos\theta, \sin\theta, 0), \quad X_2(\theta) = (0, 0, 1)$$

a cui si possono associare le derivate direzionali:

$$(2.8) \quad X_1 = \cos\theta \partial_x + \sin\theta \partial_y, \quad X_2(\theta) = \partial_{\theta}.$$

In [5] viene mostrato come un processo di integrazione eseguito su questi piani bidimensionali copre tutto lo spazio tangente tridimensionale di $\mathbb{R}^2 \times S^1$: in questo modo, a partire da X_1 e X_2 si ottiene uno spazio tangente tridimensionale in ogni punto. Questa proprietà è fondamentale per definire le connessioni tra diversi punti da ognuno dei quali, attraverso un processo di integrazione, si possono formare le cosiddette curve integrali.

I campi di associazione, mostrati in figura 2.5, possono essere interpretati come una famiglia di curve integrali dei campi vettoriali X_1 e X_2 , a partire dal punto (x, y, θ) :

$$(2.9) \quad \gamma'(t) = X_1(\gamma) + kX_2(\gamma), \quad \gamma(0) = (x_0, y_0, \theta_0)$$

ottenuti variando il parametro $k \in \mathbb{R}$, che rappresenta il coefficiente di curvatura della proiezione di γ sul piano (x, y) .

Considerando uno stimolo rappresentato dalla curva retinica:

$$\tilde{\gamma} = (x(t), y(t))$$

questa viene liftata a una curva corticale γ in $\mathbb{R}^2 \times S^1$:

$$(2.10) \quad \gamma(t) = (x(t), y(t), \theta(t))$$

tramite il metodo di soppressione dei non massimi, operato dalla corteccia per selezionare l'orientazione dello stimolo che garantisce la risposta massima in quel punto. La curva riportata in (2.9) soddisfa l'equazione (2.8), considerando le condizioni riportate in (2.7) e (2.6).

La posizione e l'orientazione di un punto in $\mathbb{R}^2 \times S^1$ possono essere scritte nel sistema di equazioni differenziali:

$$(2.11) \quad \left\{ \begin{array}{l} x' = \cos \theta \\ y' = \sin \theta \\ \theta' = k \end{array} \right.$$

e quindi γ' è parametrizzata da:

$$(2.12) \quad \gamma' = (x'(t), y'(t), \theta'(t))$$

dove x' e y' rappresentano cambiamenti in posizione, θ' in orientazione e con $k \in \mathbb{R}$ è un valore costante preso in \mathbb{R} .

Per ottenere le curve integrali, è stato discretizzato il sistema di equazioni differenziali (2.11) nel modo seguente:

$$(2.13) \quad \left\{ \begin{array}{l} \frac{x_{i+1}-x_i}{\Delta x} = \cos \theta \\ \frac{y_{i+1}-y_i}{\Delta y} = \sin \theta \\ \frac{\theta_{i+1}-\theta_i}{\Delta \theta} = k \end{array} \right.$$

Queste equazioni, caratterizzate da un valore di k costante preso in \mathbb{R} , sono in grado di modellare i campi di associazione di Field, Heyes e Hess presentati nel capitolo precedente, come mostra la figura 2.8.

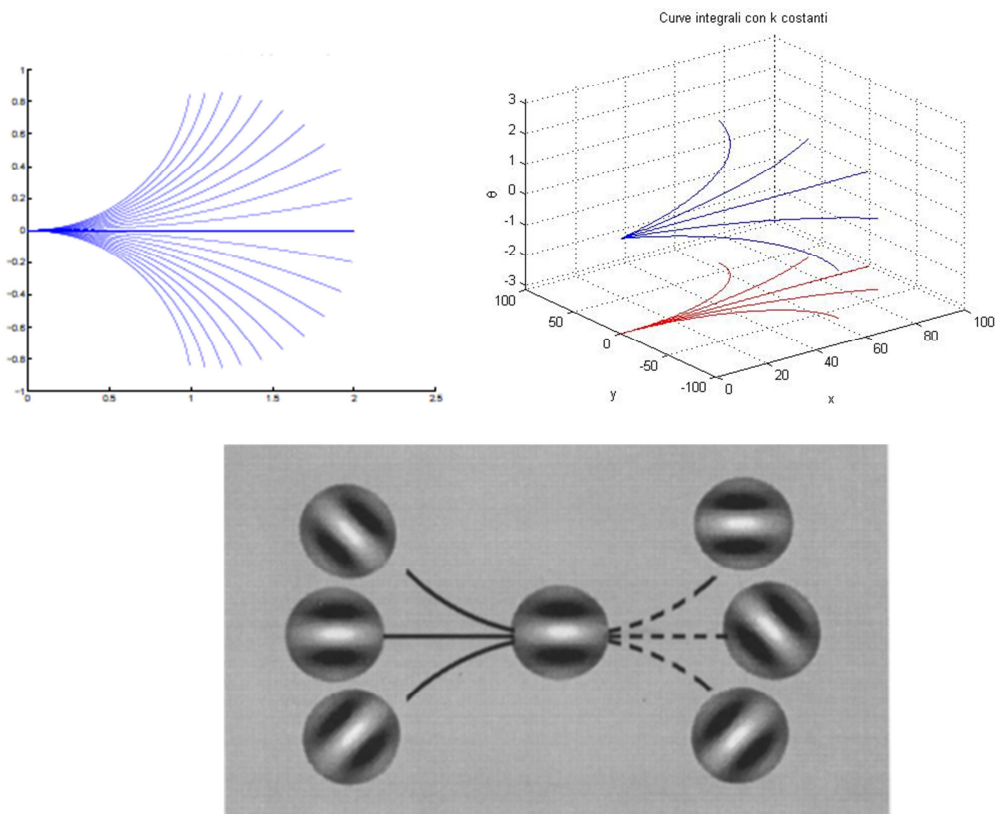


Figura 2.8:

A sinistra è presentata la proiezione di curve integrali nel piano immagine, al variare del parametro k . A destra, un esempio di curve integrali al variare di k lungo ogni percorso; in blu sono rappresentate le curve nelle tre dimensioni (x, y, θ) , in rosso la proiezione delle stesse nel piano immagine

(x, y) . Nella figura in basso, si può vedere qualitativamente come queste curve orizzontali possano essere dei buoni modelli per i campi di associazione di Field, Heyes, Hess.

Il modello dei campi di associazione di Citti e Sarti mostra quindi l'architettura funzionale della corteccia visiva descritta in termini algebrici, interpretando l'azione geometrica dei profili recettori come un 'lifting' delle curve di livello nello spazio $\mathbb{R}^2 \times S^1$. In particolare, questo approccio permette l'introduzione dell'orientazione come terza dimensione e di modellare la fenomenologia dei campi di associazione, risolvendo un sistema di equazioni differenziali.

Diverse ipotesi sono state presentate per comprendere come cellule appartenenti alla corteccia visiva primaria siano in grado di realizzare simili campi di associazione, ed in particolare, da dove queste possano prendere le informazioni necessarie per farlo. Verrà illustrato nel prossimo capitolo lo studio della connettività corticale tra neuroni sulla corteccia visiva primaria e verrà proposto un modello di connettività attraverso l'inserimento di un percorso stocastico nel sistema di equazioni differenziali (2.12).

Capitolo 3

Connettività corticale tra neuroni

Nel capitolo precedente è stato illustrato lo studio di Field, Hayes e Hess [9], i quali hanno descritto, attraverso esperimenti psicofisiologici, dei campi di associazione per stimoli orientati, che definiscono quali caratteristiche due segmenti devono avere per poter essere associati allo stesso gruppo percettivo. Ad esempio, è stato mostrato come l'abilità degli osservatori a riconoscere e estrarre contorni composti da piccoli segmenti orientati immersi in un background, dipenda sia dall'orientazione che dalla posizione degli elementi stessi. Inoltre i soggetti, sottoposti a diversi test, riescono a discriminare con maggiore facilità contorni costituiti da segmenti multipli quando questi sono allineati al percorso, rispetto a quando vi sono ortogonali. Grazie alle moderne tecniche di neuroimaging che consentono una mappatura funzionale dettagliata della corteccia visiva, è possibile giustificare su basi neurofisiologiche questi fenomeni percettivi. Ad esempio è stato osservato come neuroni sensibili ad una determinata orientazione siano preferenzialmente connessi con neuroni aventi selettività in posizione e orientazione coerenti con le regole di prossimità e buona continuazione [4]. La percezione della continuità dipende fortemente da molti aspetti che riguardano le connessioni orizzontali, le quali verranno illustrate nello specifico nella prossima sezione.

3.1 Connessioni orizzontali in corteccia

Le connessioni orizzontali sono originate dai neuroni piramidali negli strati 2/3 della corteccia visiva, si estendono per qualche millimetro parallelamente alla superficie corticale e formano terminazioni fortemente irregolari. Alcuni studi [12][23] mostravano come le connessioni orizzontali

legassero tra loro colonne di neuroni aventi simili risposte caratteristiche, come ad esempio l'orientazione. La relazione tra questo tipo di connessioni e un altro aspetto fondamentale dell'organizzazione corticale, cioè la mappa retinotopica del campo visivo, era però meno chiara. Lo studio di Bosking et al. [4] sulla corteccia striata di tupaïdi ha portato maggiore chiarezza sulle proprietà di queste connessioni, combinando imaging ottico di segnali intrinseci con piccole iniezioni di biocitina in corteccia, per quantificare la disposizione modulare e assiale delle connessioni stabilite tra neuroni.

La tecnica dell'imaging ottico utilizza diverse immagini per la localizzazione spaziale, per andare cioè a identificare aree della corteccia che rispondono principalmente a stimoli aventi una determinata orientazione nello spazio visivo. Attraverso questa tecnica, è possibile ottenere la mappa di orientazioni preferenziali, dove i colori sono usati per rappresentare l'orientazione preferenziale in ogni sito, come mostrato in figura 3.1(a). In molte regioni della mappa, chiamate "pinwheels" (girandole) è possibile notare un cambiamento nell'orientazione, campionando attorno ad un punto o singolarità. Infatti, come illustrato nel capitolo 1, la struttura della corteccia visiva presenta un'architettura funzionale organizzata in colonne, che raggruppano cellule sensibili alla stessa orientazione. Queste colonne sono organizzate nello spazio secondo una variazione di orientazione minima e continua, interrotta solo da strutture chiamate "pinwheels", in cui ogni colore rappresenta l'orientazione dello stimolo che attiva maggiormente ogni regione della corteccia. Esempi di due "pinwheels" estratti dalla mappa di figura 3.1(a) sono mostrati ingranditi in 3.1(b).

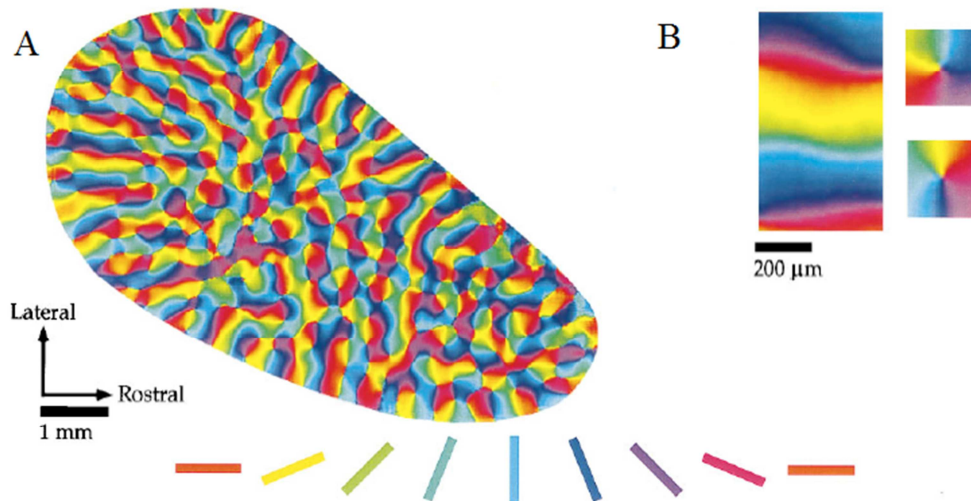


Figura 3.1:

(a): Mappa di orientazione preferenziale, ottenuta dalla somma vettoriale dei dati per ogni angolo. Il colore è codificato dai segmenti riportati sotto alla figura.

(b): Ingrandimento di una porzione della mappa riportata in (a), che dimostra sia la presenza di una zona lineare (riportata a sinistra) che della struttura a “pinwheels” (a destra).

Attraverso tecniche di imaging ottico e esperimenti di iniezioni di biocitina in corteccia è stato possibile valutare l’andamento delle connessioni orizzontali.

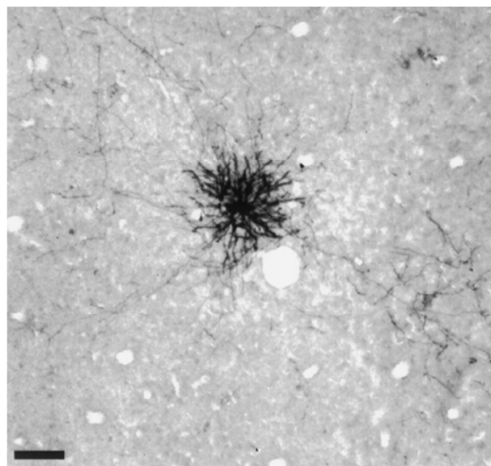


Figura 3.2:

Esempio di terminale sinaptico “bouton”. L’immagine mostra il sito di iniezione di biocitina, grazie alla quale le cellule possono essere identificate.

L’iniezione di biocitina risulta coinvolgere un basso numero di neuroni (da 12 a 65), confinati entro siti di circa 200 μm di diametro. Questi presentano il caratteristico terminale sinaptico chiamato “bouton” (figura 3.2), per la sua forma arrotondata, di dimensioni di circa 400 μm x 250 μm [4].

Per valutare al meglio l’intervallo di orientazioni principali dei siti collegati da connessioni orizzontali si esamina la distribuzione di terminali in una mappa di orientazione preferenziale colorata, come avviene in figura 3.3. Oltre all’area immediatamente adiacente al sito di iniezione, i terminali vengono raggruppati in regioni che hanno orientazioni simili a quelle del sito di iniezione.

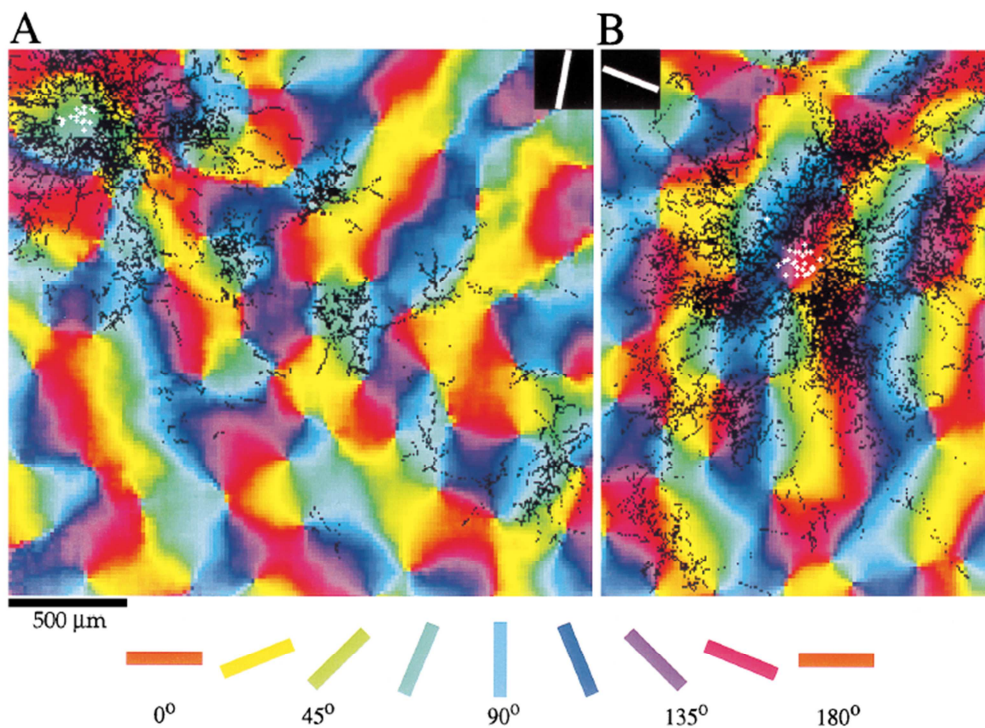


Figura 3.3:

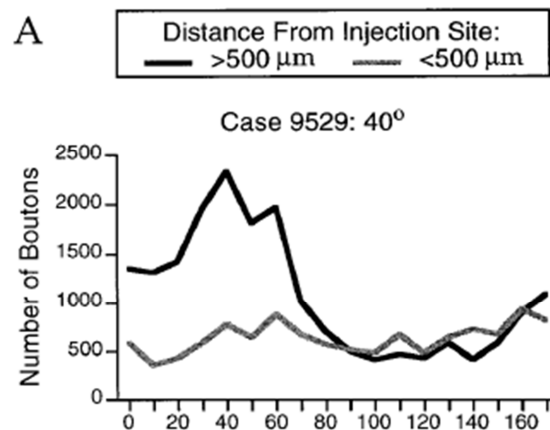
Distribuzione di terminali sinaptici “boutons” in una mappa di orientazione preferenziale in due casi:

(a): iniezione in un sito con orientazione preferita di 80° . I simboli bianchi indicano la disposizione delle cellule che assorbono biocitina. I “boutons”, rappresentati in nero, sono localizzati nei siti vicini a quello di iniezione e soprattutto in quelli aventi la stessa orientazione del sito di iniezione posti a grandi distanze

(b): iniezione in un sito con orientazione preferenziale di 160° .

In basso è riportata la scala di codifica dei colori in funzione dell'orientazione.

Per quantificare la relazione tra la distribuzione di terminali e la mappa di orientazione, sono state realizzate curve che mostrano il numero di “boutons” che si sovrappongono a una particolare valore di orientazione nella mappa. Come mostrato in figura 3.4, in ogni caso esaminato, terminali trovati a una distanza superiore a $500\ \mu\text{m}$ dal sito di iniezione mostrano un picco in corrispondenza dell'orientazione preferenziale del sito.



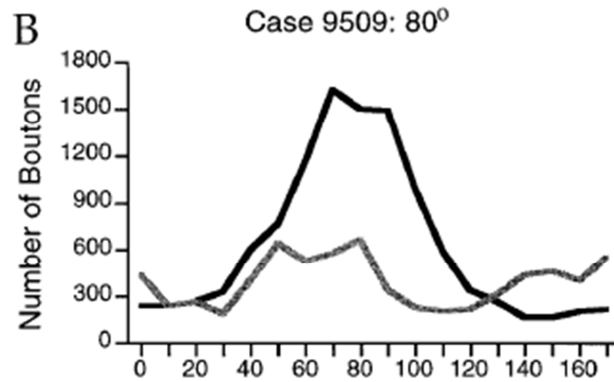


Figura 3.4:

Analisi quantitativa della distribuzione di terminali. Il numero di “boutons” trovati a distanza $< 500 \mu\text{m}$ (curve grigie) è mostrato separatamente da quelli trovati a distanza maggiore (curve nere). In ogni caso, per distanze superiori a $500 \mu\text{m}$, si ha un picco in corrispondenza dell’orientazione principale del sito di iniezione.

Le connessioni orizzontali a lungo raggio legano quindi neuroni aventi simile orientazione, inoltre viene mostrato un nuovo aspetto: l’anisotropia dell’orientazione a grandi distanze. Infatti la mappa dello spazio visivo è largamente isotropica vicino al centro di V1 e quindi le iniezioni si fanno proprio in quest’area, ma le connessioni orizzontali in corteccia si estendono per grandi distanze e distribuiscono i terminali sinaptici lungo assi della mappa del campo visivo che corrispondono a orientazioni preferenziali del sito di iniezione.

L’iniezione di biocitina nello strato 2/3 comporta una distribuzione di terminali sinaptici allungati lungo la superficie corticale. L’asse di elongazione varia da caso a caso ma è collegato all’orientazione preferenziale del sito di iniezione. Questa relazione può essere apprezzata esaminando la distribuzione di “boutons” nella figura 3.5: come mostrato in 3.5(a) un’iniezione in un sito con orientazione quasi verticale comporta una distribuzione di terminali allungati e disposti parallelamente ai bordi di V1. Diversamente, un’iniezione in un sito con orientazione quasi orizzontale da

una distribuzione di terminali perpendicolari al bordo stesso. In ogni caso si ha comunque una relazione simile: i neuroni degli strati 2/3 della corteccia danno luogo a connessioni orizzontali che si estendono per grandi distanze, lungo l'asse della mappa del campo visivo che corrisponde all'orientazione principale.

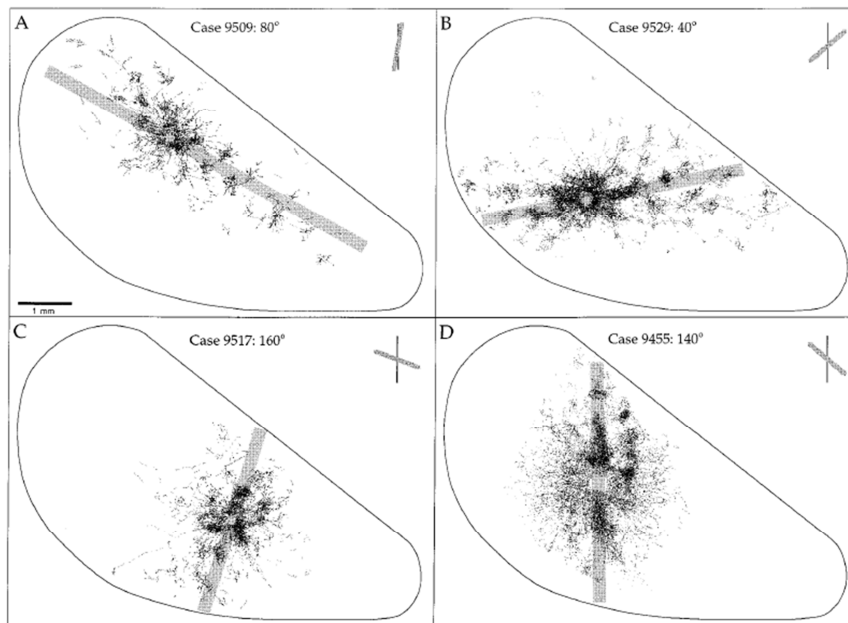


Figura 3.5:

Distribuzione di terminali sinaptici in quattro casi. L'orientazione principale per ogni esempio è mostrata in alto a destra. L'asse in corteccia corrispondente all'orientazione preferenziale è indicato dal rettangolo grigio presente in ogni distribuzione. Ogni punto indica un "bouton". Da notare l'alta densità di distribuzione di terminali vicino al sito di iniezione e una più irregolare trovata a grandi distanze. In ogni caso la distribuzione è allungata lungo l'asse che corrisponde all'orientazione preferenziale del sito di iniezione.

La distribuzione delle connessioni è legata non solo dal numero di terminali evidenziati ma anche dalla distanza di questi dal sito di iniezione. Infatti i "boutons" sono stati trovati anche a grandi distanze da questo,

lungo l'asse della mappa della campo visivo che corrisponde all'orientazione principale.

I risultati di questo studio mostrano come le connessioni orizzontali riescano a collegare neuroni aventi simile orientazione, ed inoltre si estendono per grandi distanze e distribuiscono terminali lungo l'asse che corrisponde all'orientazione preferenziale del sito di iniezione. Tuttavia i terminali sinaptici vicini a questo sito sono localizzati in regioni della mappa di orientazione che includono un intervallo esteso di valori. È presente perciò una forte anisotropia nelle connessioni orizzontali, che varia da sito a sito ed è collegata alla risposta dell'orientazione dei neuroni corticali.

I neuroni degli strati 2/3 della corteccia ricevono quindi input da altri i cui campi recettivi sono co-orientati (aventi cioè simile orientazione) e co-assiali (disposti lungo un asse nello spazio visivo che corrisponde all'orientazione principale). Questa relazione mostra come le connessioni orizzontali siano il maggior contributo alla selettività dell'orientazione dei neuroni negli strati 2/3. Per esempio, un neurone che risponde al meglio ad uno stimolo verticale si comporta così perché riceve input da una rete di neuroni degli strati 2/3 i cui campi recettivi sono allineati lungo l'asse verticale dello spazio visivo.

Le connessioni orizzontali esercitano una maggior influenza nelle regioni dello spazio visivo che si trovano lungo gli assi di orientazione principale dei neuroni. Infine, anche la percezione della continuità nel percorso visivo dipende dagli effetti delle connessioni orizzontali a lungo raggio. Per esempio, l'abilità di un osservatore ad estrarre contorni composti da piccoli segmenti orientati immersi in un background dipende sia dall'orientazione che dalla posizione degli elementi stessi: i soggetti riescono a identificare al meglio contorni composti da segmenti multipli quando allineati al percorso piuttosto che quando sono ortogonali allo stesso.

Si può quindi sottolineare l'importanza delle connessioni orizzontali nella creazione di un campo di associazione, nel quale devono essere soddisfatte

determinate condizioni di orientazione e posizione. Nello scorso capitolo è stato illustrato il modello di campi di associazione nello spazio $\mathbb{R}^2 \times \mathcal{S}^1$ introdotto da Citti e Sarti, che fornisce una giustificazione del completamento percettivo sulla base della funzionalità della corteccia visiva primaria (V1). Attraverso questo modello è stato possibile modellare la connettività cellulare.

3.2 Costruzione di kernel stocastici

Lo studio di Field, Hayes e Hess [9] ha introdotto la nozione di campo di associazione, come un percorso di integrazione di informazioni tra immagini che soddisfa il principio di buona continuazione. Il ruolo principale di questo meccanismo si ha nella percezione e integrazione dei contorni. Quasi contemporaneamente, Mumford [25] ha proposto un approccio differente per descriverli, utilizzando un processo stocastico e definendo curve con curvatura variabile in ogni punto. Questo ha prodotto delle distribuzioni di probabilità nello spazio $\mathbb{R}^2 \times \mathcal{S}^1$.

Williams e Jacobs [16] hanno utilizzato tale processo stocastico per implementare un meccanismo di completamento e hanno interpretato il kernel ottenuto come un tensore che rappresenta connessioni geometriche nello spazio, associato alla rappresentazione neurale delle immagini dovuta alle cellule semplici. L'assunzione fondamentale è che i percorsi seguiti da particelle lungo un percorso casuale in un reticolo, con posizioni e orientazioni discrete, possono essere utilizzati per modellare la distribuzione di probabilità nel completamento visivo. L'attività di neuroni con determinati campi recettivi può essere quindi interpretata come una distribuzione di probabilità, che descrive la possibile posizione assunta da una particella. Si definisce così un campo di completamento stocastico come la probabilità che una particella, inizialmente in uno stato, passi attraverso un percorso casuale, nello stato finale: il completamento visivo viene perciò modellato come un 'random walk'. Questo moto casuale segue

le leggi della Gestalt di prossimità e buona continuazione e le particelle sono descritte sia da orientazione che posizione.

Per ottenere delle densità di probabilità e definire così la connessione tra cellule semplici in V1 è stato introdotto, nel capitolo precedente, il modello dei campi di associazione di Citti e Sarti. Questo mostra l'architettura funzionale della corteccia visiva descritta in termini algebrici, interpretando l'azione geometrica dei profili recettori come un 'lifting' delle curve di livello nello spazio $\mathbb{R}^2 \times S^1$.

Questo ha permesso di modellare la fenomenologia dei campi di associazione, risolvendo un sistema di equazioni differenziali (2.14), con $k \in \mathbb{R}$ [5]. La soluzione a questo sistema rappresenta una famiglia di curve integrali (figura 2.8) che può essere utilizzata per modellare i campi di associazione e può dare informazioni sulla loro struttura. Per modellare la connettività corticale tra neuroni nella V1, si propone un modello attraverso l'inserimento di un processo stocastico al sistema (2.14).

Si consideri dunque il sistema di equazioni differenziali stocastiche:

$$(3.1) \quad \begin{cases} x' = \cos \theta \\ y' = \sin \theta \\ \theta' = k \end{cases}$$

dove x' e y' rappresentano cambiamenti in posizione, θ' in orientazione e k è una variabile casuale, derivante da una distribuzione Gaussiana, con media 0 e varianza σ^2 . L'effetto è che le particelle tendono a seguire delle linee dritte ma, nel tempo, deviano a sinistra o destra secondo un termine proporzionale a σ^2 . Quando $\sigma^2 = 0$ il moto è completamente deterministico e le particelle non deviano mai dal percorso rettilineo, quando invece $\sigma^2 = \infty$ (quindi per k variabile) il moto è totalmente casuale e il processo stocastico diventa un 'random walk': in questo caso si parla di processo diffusivo, con σ^2 che è la costante di diffusione [16].

La soluzione a questo sistema di equazioni differenziali rappresenta una famiglia di curve integrali che può essere utilizzata per modellare le connettività orizzontali e può dare informazioni sulla loro struttura.

Per ottenere le curve integrali di figura 3.6, è stato discretizzato il sistema di equazioni differenziali nel modo seguente:

$$(3.2) \quad \left\{ \begin{array}{l} \frac{x_{i+1}-x_i}{\Delta x} = \cos \theta \\ \frac{y_{i+1}-y_i}{\Delta y} = \sin \theta \\ \frac{\theta_{i+1}-\theta_i}{\Delta \theta} = \delta_\theta \end{array} \right.$$

Per costruire un nucleo differenziale stocastico rappresentativo dei campi di associazione, occorre quindi considerare al posto di k un moto browniano o processo stocastico: si ha così un sistema di equazioni differenziali stocastiche e risolvendolo per una serie di cammini e integrando per la variabile d'arco si ottiene proprio il nucleo cercato.

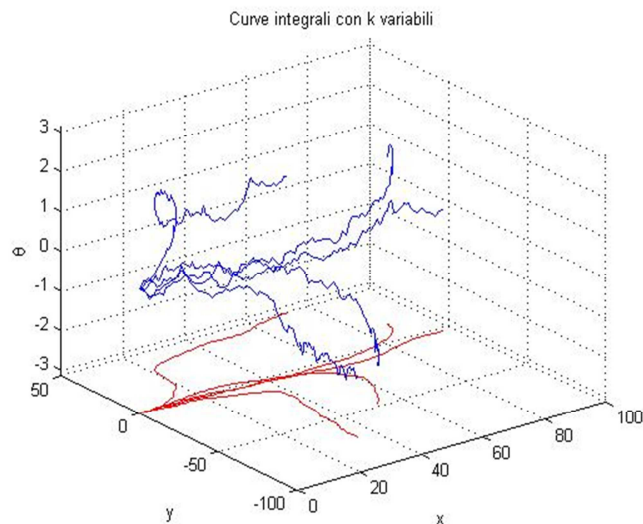


Figura 3.6:

Esempio di curve integrali al variare di k lungo ogni percorso, dove k rappresenta un processo stocastico. In blu le curve nelle tre dimensioni (x, y, θ) , in rosso la proiezione delle stesse nel piano immagine (x, y) .

L'attività dei neuroni in V1 può essere quindi modellata da una funzione densità di probabilità definita in $\mathbb{R}^2 \times S^1$, dove si ha un insieme di orientazioni locali associato ad ogni punto del piano immagine.

Questo ha permesso di modellare la connettività cellulare, risolvendo un sistema di equazioni differenziali stocastiche (3.2), con k processo stocastico [5]. Si sono ottenuti così un nucleo e la densità di probabilità associata, che modella l'attività di neuroni corticali. Per modellare la connettività orizzontali tra cellule semplici, è stato sviluppato un metodo per la costruzione di kernel stocastici tridimensionali.

Questi nuclei sono dipendenti da diversi parametri, come le loro dimensioni, il punto iniziale, l'asse di orientazione, ma quelli che risultano più influenti sono: il numero di percorsi, il numero di passi utilizzati in ognuno dei percorsi e la costante di diffusione σ^2 .

Ogni nucleo permette il calcolo della distribuzione di probabilità che un punto qualunque sia connesso con quello iniziale. Nella sua costruzione si sono utilizzati alcuni accorgimenti: si considera un kernel di dimensioni dispari, per avere una posizione iniziale univoca, non ambigua e per evitare l'asimmetria nei valori, inoltre, per poter fare successivamente grouping, occorrono affinità e nucleo completo. Per fare ciò, sono stati valutate entrambe le direzioni delle curve integrali, lungo la stessa orientazione. Vediamo ora come cambia il nucleo al variare dei parametri sopraelencati.

Il numero di step, o numero di passi in un percorso, rappresenta un indice di "lontananza" per il kernel: al diminuire di questa variabile il nucleo non risulta definito oltre un certo raggio, come mostrato in figura 3.7.

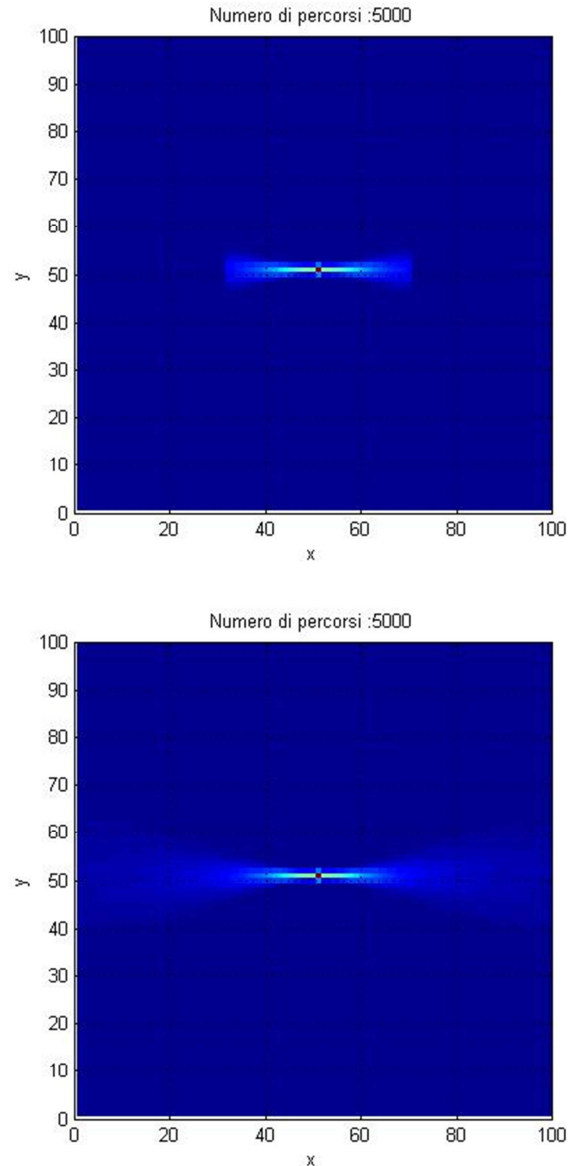


Figura 3.7:

Esempi di nuclei con numero di step variabile e pari a: 20 per l'immagine in alto, 100 per quella in basso. Come si nota, il secondo caso presenta un nucleo con raggio più ampio. Le immagini sono state calcolate con 5000 percorsi, costante di diffusione $\sigma^2=0.05$ e asse del nucleo orizzontale.

Il secondo parametro di interesse è il numero di percorsi, che influisce sull'omogeneità del kernel. Come mostrato in figura 3.8, per un basso numero di percorsi il kernel risulta estremamente discontinuo, a mano a mano che i percorsi aumentano il nucleo diventa più ordinato e uniforme.

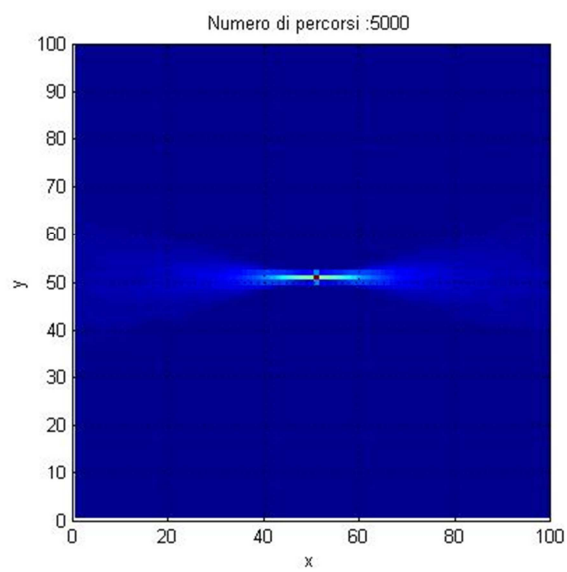
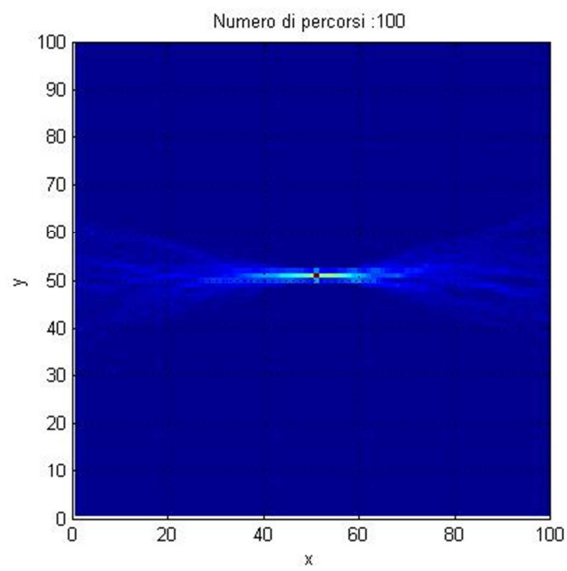
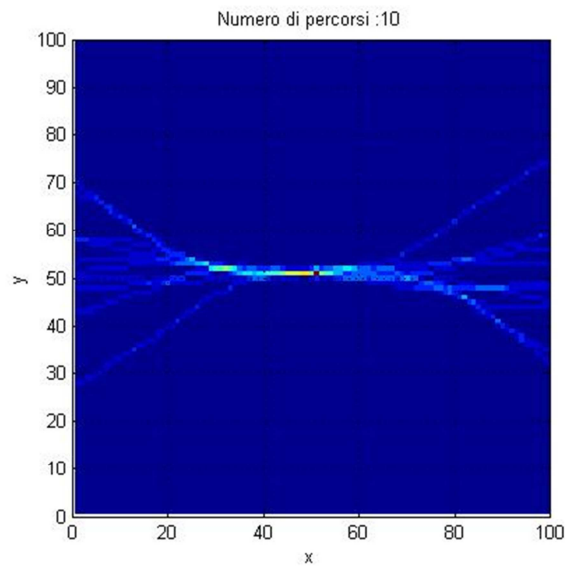


Figura 3.8:

Esempi di nuclei con numero di percorsi variabile e pari a: 10, 100, 5000 (dall'alto verso il basso). All'aumentare del parametro il kernel diventa più omogeneo. Le immagini sono state calcolate con 200 step, costante di diffusione $\sigma^2=0.05$ e asse del nucleo orizzontale.

È inoltre possibile determinare l'errore che si commette calcolando il nucleo con un diverso numero di percorsi. La variazione casuale fa sì che una misura effettuata su un campione non fornisca un valore identico a quello ottenibile misurando l'intera popolazione: c'è sempre un certo errore, che viene detto errore campionario. Questo errore deriva semplicemente dal fatto che stiamo osservando soltanto una parte della popolazione.

All'aumentare perciò del numero di percorsi l'errore diminuisce, con andamento esponenziale decrescente: idealmente, al tendere all'infinito dei percorsi, l'errore tende a 0.

L'ultima variabile presa in considerazione per lo studio del nucleo è la costante di diffusione σ^2 . Questa rappresenta un indice di "ampiezza" del nucleo: all'aumentare di σ^2 il kernel diventa più largo, diminuendo σ^2 invece il kernel degenera in una retta (figura 3.9; verrà successivamente illustrato che un basso valore di diffusione comporta una buona affinità per segmenti allineati).

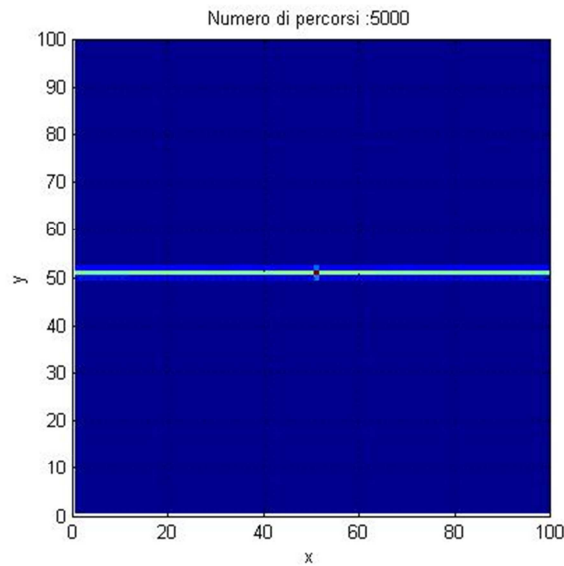


Figura 3.9:

Esempio di nucleo rettilineo con asse orizzontale, caratterizzato da 5000 percorsi, 200 step e $\sigma^2=0.001$.

Nelle figure 3.7, 3.8, 3.9 sono riportati nuclei con assi orizzontali ($\theta = 0$), è possibile però fare gli stessi ragionamenti con diverse orientazioni. Un esempio è mostrato in figura 3.10: in rosso è rappresentato il nucleo orizzontale, in verde il kernel con asse a $\theta = \pi/4$, che risulta ruotato e traslato nella fibra di orientazione.

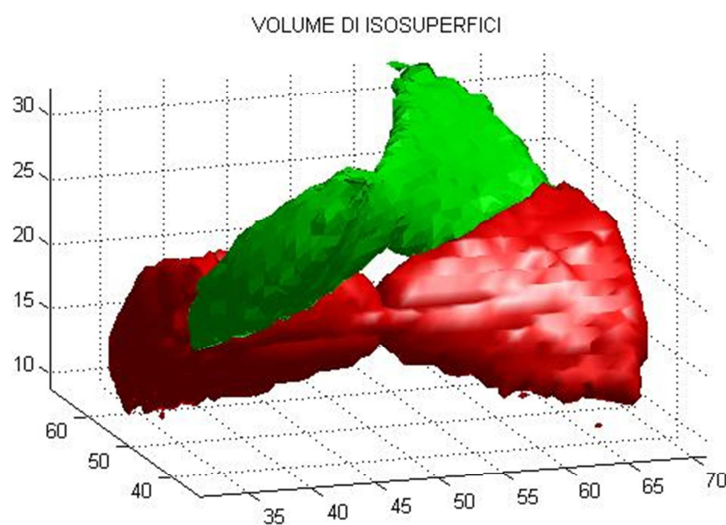


Figura 3.10:

Rappresentazione di un nucleo con asse orizzontale (rosso) e uno con asse a 45°(verde); i nuclei sono stati calcolati con 5000 percorsi, 20 step e $\sigma^2=0.05$.

I nuclei sopra rappresentati sono delle densità di probabilità di connessione tra cellule, a cui è possibile collegare direttamente il concetto di affinità tra stimoli visivi.

Come verrà illustrato nel prossimo capitolo, a partire da un numero prefissato di punti orientati è possibile infatti calcolare l'affinità tra stimoli co-circolari, attraverso l'utilizzo del kernel stocastico e la costruzione della corrispondente matrice. Come per il nucleo, verrà successivamente mostrato come quest'ultima cambi al variare degli stessi parametri e proprio sulla costruzione di determinate matrici si sono basati diversi metodi di riduzione di dimensionalità. La fenomenologia del grouping visivo è, di fatto, il risultato di un procedimento di riduzione di dimensionalità. Attraverso l'analisi spettrale di matrici di affinità calcolate utilizzando i modelli geometrici di Citti-Sarti, verrà rappresentata la possibilità di poter riprodurre il fenomeno percettivo di grouping nello spazio $\mathbb{R}^2 \times S^1$, mostrando i risultati e gli esempi applicativi sviluppati.

Capitolo 4

Applicazione del modello di connettività: il raggruppamento percettivo

Nel precedente capitolo è stato introdotto il concetto di affinità tra stimoli visivi, che può essere utilizzato per modellare la fenomenologia del grouping. Per fare questo si utilizzano tecniche di analisi spettrale di matrici di affinità, quali ad esempio tecniche di riduzione di dimensionalità. Queste occupano una posizione centrale in molti campi, quali la teoria dell'informazione e del campionamento e la statistica. In sostanza, lo scopo è quello di cambiare la rappresentazione dei dati, i quali originariamente coinvolgono un gran numero di variabili, in una descrizione degli stessi a dimensione inferiore, usando un minor numero di coordinate. La nuova rappresentazione deve descrivere i dati in maniera fedele, mantenendo le quantità di interesse, come al esempio una nozione di distanza tra gli stessi. Analogamente al problema di riduzione di dimensionalità, vi è quello di trovare strutture significative nell'insieme dei dati, con lo scopo di estrarre caratteristiche rilevanti per capire il fenomeno che li ha generati. Per raggiungere gli scopi suddetti, sono state sviluppate numerose tecniche che utilizzano algoritmi basati su grafi, i quali offrono un vantaggioso compromesso tra semplicità, interpretabilità e la loro abilità di rappresentare complesse relazioni tra i dati. I grafi pesati sono solitamente utilizzati per descrivere le interazioni tra punti. In molte situazioni, ogni campione è rappresentato da un insieme di attributi numerici e, in questo caso, la condizione affinché due nodi siano connessi (e di conseguenza la forza di questa connessione) è basata su una definizione di vicinanza tra i punti nello spazio dei dati. In particolare, nei campi di classificazione e clustering, i grafi sono risultati essere molto efficienti nel trovare strutture rilevanti in geometrie complesse [6]. Verrà illustrato nella prossima sezione proprio il concetto di clustering come tecnica di analisi dei dati che prova a

identificare gruppi aventi elementi affini, ed in particolare verranno poi descritti gli algoritmi di Spectral clustering e il K-means, mostrandone connessioni e divergenze tra metodi. Nelle prossime sezioni verrà illustrata la costruzione della matrice di affinità al variare dei parametri del nucleo, riportando diversi casi. Infine, si mostreranno esempi applicativi, quali l'articolazione figura/sfondo e l'estrazione di figure con tecniche di riduzione di dimensionalità.

4.1 Grouping e clustering

Con il termine grouping si intende l'unione di alcuni elementi di un set di dati in un unico gruppo o unità. Questo concetto è intimamente connesso a quello di clustering, attraverso il quale si suddivide un certo insieme di dati in un numero noto di partizioni. Al termine cluster si attribuisce il significato di "gruppo compatto", dove per compatto si intende che, data una qualche misura di distanza, gli elementi del gruppo sono vicini tra loro e lontani dagli altri gruppi. Per classificare un set di dati, bisogna quindi introdurre una nozione di distanza tra ogni elemento, una funzione che ci dia anche una misura della loro similarità. Una distanza che si utilizza spesso è quella euclidea: dati due punti (x, y) definiti in un spazio \mathbb{R}^n , questa è definita da:

$$(4.1) \quad d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Tra gli algoritmi per fare raggruppamenti, il K-means è tra i più semplici e rappresentativi. Sia dato il training set e il numero di classi k in cui lo si vuole partizionare. Si istanziano k centroidi, o prototipi, che rappresentano una classe, inizialmente scelti casualmente nello spazio contenente il training set; in seguito la loro posizione viene aggiornata in modo tale da minimizzare una funzione energia:

$$(4.2) \quad V = \sum_{i=1}^k \sum_j d(x_j - \mu_i)^2$$

dove μ_i è l' i -esimo centroide e d solitamente la distanza euclidea. I passi del I° algoritmo presentato sono:

input: training set, numero k di classi e misura di distanza.

1. Inizializzare casualmente k prototipi nello spazio contenente i dati.
2. Assegnare ogni punto al cluster del centroide più vicino.
3. Ricalcolare ogni prototipo in modo che si trovi nel baricentro dei dati della propria classe.
4. Ritornare al punto 2 a meno che non venga soddisfatta una condizione di stop (che può essere, ad esempio, l'invarianza della posizione dei prototipi rispetto al ciclo precedente).

output: cluster o i prototipi.

Un esempio di clustering attraverso l'algoritmo del K-means è riportato in figura:

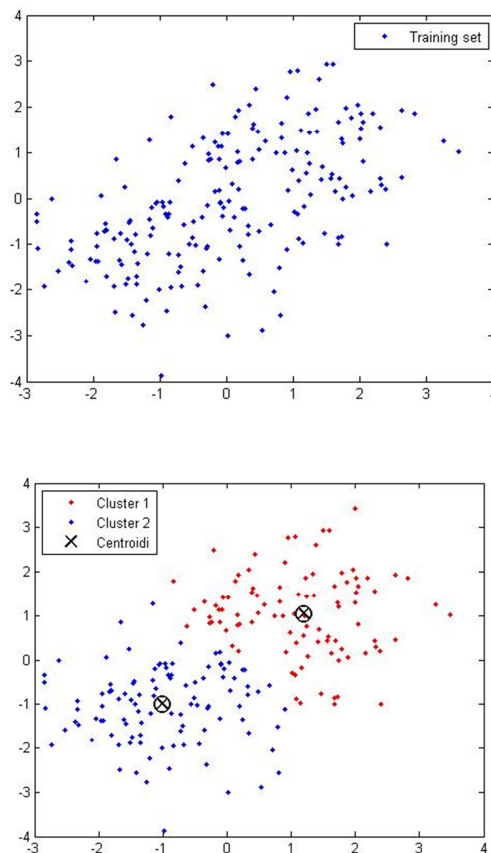


Figura 4.1: Esempio di creazione di due cluster partendo da dati random.

La funzione di energia (4.2) che viene minimizzata dal K-means ha un minimo globale ma anche dei minimi locali. Uno dei problemi di questa tecnica è proprio il fatto di cadere spesso in uno di questi minimi locali. Per risolverlo in parte, l'algoritmo viene solitamente eseguito più volte e vengono confrontati i risultati. Un altro svantaggio del K-means è che il numero di partizioni è difficile da stabilire a priori: l'algoritmo non è in grado di definire il valore ottimale di k . Un vantaggio, oltre alla facilità di calcolo, è invece rappresentato dalla bassa complessità computazionale del K-means che è in media $O(nk)$, dove n è la cardinalità dell'insieme dei dati e k è il numero di centroidi stabilito a priori [22].

4.2 Spectral Clustering

In questa sezione verranno presentate le tecniche per estrarre informazioni di grouping visivo, attraverso l'analisi dello spettro di una matrice di affinità.

Negli ultimi anni, diverse tecniche si sono diffuse come valide alternative ad algoritmi tradizionali, come il K-means, rispetto ai quali risultano essere molto più vantaggiose e performanti, in particolare tecniche di analisi spettrale. Tra i motivi, la semplicità dell'implementazione e il fatto che questi algoritmi possano essere risolti con semplici metodi di algebra lineare. La teoria che sta alla base di queste tecniche è la teoria dei grafi [29]. Dato un insieme di punti x_1, \dots, x_n e una qualche nozione di similarità $w_{i,j} \geq 0$ definita tra coppie di questi, lo scopo del clustering è quello di dividere il set di dati in gruppi in modo tale che punti appartenenti allo stesso gruppo siano simili e punti in gruppi differenti siano diversi tra loro. Un modo per rappresentare i dati è nella forma dei grafi di affinità, dove ogni vertice v_i rappresenta un punto x_i ; due vertici risultano connessi se

l'affinità $w_{i,j}$ tra coppie di punti x_i e x_j è positiva o comunque maggiore di una certa soglia.

Dati dei punti v_1, \dots, v_n un grafo può essere definito come $G = (V, E)$ dove $V = (v_1, \dots, v_n)$ è l'insieme dei vertici, o nodi, del grafo ed $E = \{(v_i, v_j) \mid v_i, v_j \text{ sono connessi}\}$ è l'insieme degli archi che connettono dei vertici distinti. Si assume che nel grafo la connessione sia pesata da una misura di similarità $w_{i,j} \geq 0$: l'insieme delle connessioni può essere rappresentato da una matrice di adiacenza $W = (w_{i,j})_{i,j=1\dots n}$. Se $w_{i,j} = 0$ significa che i due vertici (v_i, v_j) non sono collegati. Se la matrice W è simmetrica, allora il grafo G che rappresenta è non direzionato e quindi il peso $w_{i,j} = w_{j,i}$; se invece la matrice W è simmetrica il grafo che rappresenta è direzionato, cioè c'è differenza nell'andare dal nodo i al nodo j e nell'andare dal nodo j al nodo i . Il grado d_i di un vertice $v_i \in V$ è definito da:

$$(4.3) \quad d_i = \sum_{j=1}^n w_{i,j}$$

Il grado di un vertice quindi non è altro che la somma dei pesi delle sue connessioni con altri vertici adiacenti. La matrice grado D è definita come la matrice diagonale contenente i gradi dei vertici come elementi non nulli.

A partire da un grafo si costruisce la matrice di affinità W attraverso vari metodi.

Grafi di ε -vicinanza: Sono considerate connesse le coppie di vertici la cui distanza euclidea è inferiore a ε , in tal caso la connessione vale 1, altrimenti 0.

Grafi k-primi vicini: Ogni vertice v_i è connesso con v_j se v_j è tra i k primi vertici vicini a v_i .

Grafi completamente connessi: Tutti i punti sono connessi tra loro. Per ottenere questo tipo di grafo o matrice si può utilizzare ad esempio una funzione Gaussiana sulla distanza: $w(x_i, x_j) = e^{(-\|x_i - x_j\|^2 / 2\sigma^2)} = e^{(-\|dist\|^2 / 2\sigma^2)}$, dove il parametro σ controlla l'ampiezza della vicinanza. Questa funzione viene utilizzata per questo tipo di grafo in quanto non si annulla mai.

Come verrà successivamente illustrato, tutti questi grafi sono utilizzati regolarmente nello spectral clustering [28]. Negli esempi applicativi analizzati si utilizzano grafi e matrici di affinità che idealmente sono completamente connessi.

La matrice di similarità W rappresenta le connessioni che caratterizzano un grafo e da questa si può costruire una matrice di fondamentale importanza nello spectral clustering, il Laplaciano. Quest'ultimo si può distinguere tra non normalizzato e normalizzato.

Il Laplaciano non normalizzato è definito da:

$$(4.4) \quad L = D - W$$

e questa matrice risulta simmetrica e semi-definita positiva, ed ha n autovalori reali positivi: $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ [28]. Sono presenti invece due matrici, collegate tra di loro, per rappresentare il Laplaciano normalizzato:

$$(4.5) \quad \begin{aligned} L_{rw} &= D^{-1}L = I - D^{-1}W \\ L_{sym} &= D^{-1/2}LD^{-1/2} = I - D^{-1/2}WD^{-1/2} \end{aligned}$$

dove L_{sym} è una matrice simmetrica mentre L_{rw} è collegata a random walk. Queste sono caratterizzate dalle seguenti proprietà:

- 1) λ è autovalore di L_{rw} con autovettore u se e solo se λ è autovalore di L_{sym} con autovettore $w = D^{-1/2}u$,
- 2) λ è autovalore di L_{rw} con autovettore u se e solo se λ e u risolvono il problema generalizzato: $Lu = \lambda Du$,
- 3) L_{sym} e L_{rw} sono simmetriche e semi-definite positive, con autovalori reali e positivi: $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$.

Una questione fondamentale correlata allo spectral clustering è quale dei tre grafi Laplaciani debba essere usato per il calcolo degli autovettori. Il primo argomento da considerare viene dalla partizione del grafo stesso. Considerando per semplicità il caso di $k=2$, il clustering ha due obiettivi diversi:

- 1) trovare una partizione in modo che punti in cluster diversi non siano simili tra loro, cioè si minimizza la similarità tra gruppi,
- 2) trovare una partizione in modo che punti nello stesso cluster siano simili tra loro, si vuole cioè massimizzare la similarità all'interno di un gruppo.

Il Laplaciano normalizzato implementa entrambi gli obiettivi riportati sopra mentre quello non normalizzato soddisfa solo il primo punto. Tra le sue due matrici, si tende ad usare L_{rw} poiché gli autovettori di L_{sym} sono moltiplicati per $D^{1/2}$ e questo può portare ad artefatti indesiderati.

Infatti, i due approcci forniscono risultati molto simili nel caso in cui i gradi dei vertici del grafo siano simili tra loro. Nel caso in cui i gradi dei vertici assumano valori abbastanza diversi i due approcci sono distinti, in particolare il Laplaciano non normalizzato dà importanza a minimizzare il taglio e a massimizzare la cardinalità per partizione, piuttosto che il volume. I due concetti, cardinalità e volume, sono scorrelati tra loro, il primo si basa sul numero di vertici, il secondo sulle connessioni tra gli stessi. Dal punto di vista dei grafi il volume è una proprietà più interessante che indica la connettività interna di una partizione; questa è una prima motivazione per preferire la definizione di Laplaciano normalizzato.

Oltre a questo argomento in [28] viene mostrata un'altra differenza riguardante i due approcci: sotto alcune condizioni il Laplaciano normalizzato porta a risultati statisticamente consistenti, al contrario del Laplaciano non-Normalizzato. Questo significa che se la cardinalità del training set tende a infinito i risultati del primo convergono ad una partizione, ma questo non vale necessariamente nel secondo caso.

Prima di definire gli algoritmi di Spectral clustering, occorre introdurre un concetto fondamentale: il taglio del grafo. Considerato il training set come se fosse un grafo pesato G , lo Spectral clustering ha l'obiettivo di 'tagliarlo' nel miglior modo possibile, separando punti in diversi gruppi in accordo alla loro similarità. Un taglio è una partizione del grafo in due insiemi disgiunti e ha un valore, chiamato cut. Se A e B sono due insiemi disgiunti il valore del taglio è la somma dei pesi delle connessioni che vengono 'tagliate':

$$(4.6) \quad cut(A, B) = \sum_{i \in A, j \in B} w_{i,j}$$

Un primo approccio al problema del clustering utilizzando un grafo è stato proprio quello di minimizzare questa quantità: questo porta però a scarsi risultati, poiché spesso viene tagliato via solo un vertice, come nell'esempio riportato in figura.

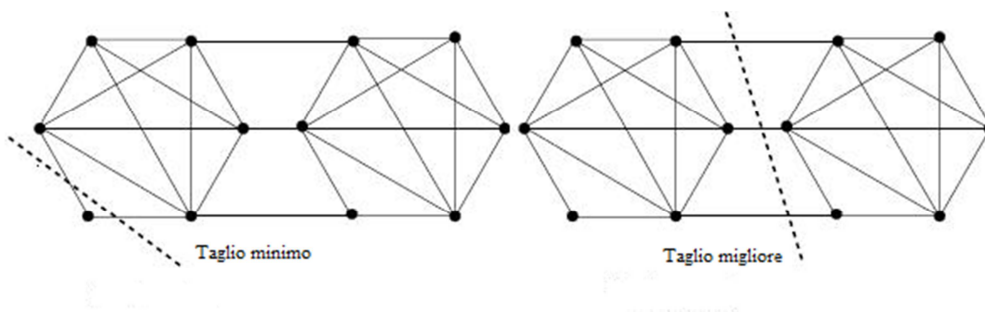


Figura 4.2: Esempio di taglio minimo e bilanciato.

Quello che rende migliore il taglio di figura 4.2 è il volume delle parti tagliate, Shi e Malik hanno sfruttato questo concetto, aggiungendo il vincolo che i due insiemi divisi siano abbastanza voluminosi [22].

Si definisce il taglio normalizzato come il valore di taglio di un grafo normalizzato rispetto ai volumi degli insiemi che ne risultano:

$$(4.7) \quad Ncut(A, B) = cut(A, B) \left(\frac{1}{Vol(A)} + \frac{1}{Vol(B)} \right)$$

Con questa formulazione minimizzare $Ncut$ ‘tagliando’ un solo vertice non è più possibile, essendo i volumi di A e B non massimizzati.

In questo senso Shi e Malik hanno dimostrato che minimizzare il taglio normalizzato conduce contemporaneamente alla massimizzazione della connessione interna degli insiemi: si massimizza così l’affinità intragruppo. Un’altra soluzione per risolvere il problema del ‘taglio’ attraverso la richiesta di trovare volumi ragionevolmente grandi, è rappresentato dal *RatioCut* [28]. Sia dato un grafo di similarità con matrice di adiacenza: $W(A, B) = \sum_{i \in A, j \in B} w_{i,j}$, sia \bar{A} il complemento di A . Per un dato numero k di sottoinsiemi, l’approccio del minimo taglio consiste nel scegliere una partizione A_1, \dots, A_k che minimizzi:

$$(4.8) \quad cut(A_1, \dots, A_k) = \frac{1}{2} \sum_{i=1}^k W(A, \bar{A}_i)$$

in modo analogo all’equazione (6). Si introduce il fattore $\frac{1}{2}$ perché altrimenti ogni arco verrebbe contato due volte. Come visto, questo problema non porta spesso a partizioni soddisfacenti. Mentre per $Ncut$ la misura di un sottoinsieme di A è pesata dal suo volume, nel *RatioCut* questa è misurata dal suo numero di vertici $|A|$:

$$(4.9) \quad RatioCut(A_1, \dots, A_k) = \frac{1}{2} \sum_{i=1}^k \frac{W(A_i, \bar{A}_i)}{|A_i|}$$

Da notare che questa funzione, come $Ncut$, assume piccoli valori se i cluster A_i non sono troppo piccoli. In particolare il minimo di $\sum_{i=1}^k \frac{1}{|A_i|}$ è raggiunto se tutti gli $|A_i|$ coincidono. Il $RatioCut$ minimizza quindi il valore di taglio sul grafo, massimizzando la cardinalità dei sottogruppi.

Entrambe queste funzioni provano a fare in modo che i cluster siano bilanciati, come misurato dal numero di vertici o dai bordi pesati, rispettivamente. Sfortunatamente, introdurre queste condizioni rende i problemi NP hard: lo spectral clustering è un modo per risolvere la versione rilassata di questi. In particolare, il rilassamento di $Ncut$ porta all'algoritmo del Normalized spectral clustering, mentre il $RatioCut$ alla versione non normalizzata [28].

Un'ulteriore linea di argomenti per spiegare lo Spectral clustering è basata sui random walks nei grafi di similarità, che rappresentano un processo stocastico che causalmente passa da un vertice a un altro. In questo, lo Spectral clustering cerca di trovare una partizione del grafo in modo che i 'cammini casuali' restino all'interno dello stesso cluster, con qualche raro salto tra i gruppi. Intuitivamente questo ha senso, in particolare se si considera insieme alla proprietà del taglio del grafo: una partizione bilanciata con un basso valore di taglio avrà anche la caratteristica che i random walks non abbiano l'opportunità di passare da un cluster a un altro. La probabilità di transizione di un salto dal vertice v_i a v_j è proporzionale al peso w_{ij} ed è data da: $p_{ij} = w_{ij}/d_i$. La matrice di transizione $P = ((p_{ij})_{i,j=1,\dots,n})$ del random walk è definita da:

$$(4.10) \quad P = D^{-1}W$$

Esiste una stretta relazione tra L_{rw} e P , considerando le equazioni riportate in (4.5) e (4.10) infatti si ha: $L_{rw} = I - P$. Come conseguenza, λ è autovalore di L_{rw} con autovettore u se e solo se $1 - \lambda$ è autovalore di P con autovettore u ; inoltre il più grande autovettore di P e il più piccolo

autovettore di L_{rw} possono essere usati per descrivere le proprietà dei cluster nel grafo. Un'equivalenza formale tra $Ncut$ e la probabilità di transizione di random walk è stata osservata da Shi e Malik [22]: minimizzare $Ncut$ porta a tagliare il grafo in modo che i 'cammini' raramente passino da A a \bar{A} e viceversa.

Verrà ora illustrato l'algoritmo di Spectral Clustering di Shi Malik: siano dati n punti x_1, \dots, x_n e la loro similarità a coppie $s_{ij} = s(x_i, x_j)$, calcolata da una funzione simmetrica e non negativa, e sia $S = (s_{ij})_{i,j=1,\dots,n}$ la corrispondente matrice [22].

II° algoritmo: Normalized Spectral clustering.

Input: matrice di similarità S e condizione di stop

- 1) Costruire un grafo di similarità in uno dei modi descritti nella scorsa sezione. Sia W la matrice di adiacenza pesata.
- 2) Calcolare il Laplaciano L .
- 3) Calcolare i primi k autovettori u_1, \dots, u_n del problema generalizzato:
$$Lu = \lambda Du.$$
- 4) Sia $U \in R^{n \times k}$ la matrice contenente i vettori u_1, \dots, u_k come colonne.
- 5) Per $i = 1, \dots, n$, sia $y_i \in R^k$ il vettore corrispondente alla i -esima riga di U .
- 6) Raggruppare i punti $(y_i)_{i=1,\dots,n}$ in R^k con l'algoritmo del k -means in cluster C_1, \dots, C_k .

Output: Cluster A_1, \dots, A_k con $A_i = \{j \mid y_j \in C_i\}$.

Da notare che quest'algoritmo lavora con gli autovettori del Laplaciano normalizzato L_{rw} , perciò si parla di Normalized Spectral clustering.

Un altro algoritmo interessante per il clustering è quello proposto da Perona Freeman [10], il quale permette di separare in una determinata regione dello spazio il 'foreground' dal 'background'. Calcolando l'approssimazione puntuale p_i della matrice di affinità S , contenente l'affinità tra coppie di punti, è possibile infatti separare la scena nei due gruppi, dove p

rappresenta una funzione che da la salienza dei punti stessi. Il III° algoritmo, chiamato “Fattorizzazione di affinità” è il seguente:

- 1) Costruire la matrice di affinità S
- 2) Sia p l'autovettore di S corrispondente al suo più grande autovalore
- 3) Si definisce ‘foreground’ F l'insieme di oggetti i i cui corrispondenti p_i sono diversi da zero (o meglio $p > \varepsilon > 0$).

Questi due algoritmi sono stati confrontati [10] attraverso esperimenti che mostrano come entrambi si comportino bene, raggiungendo un buon raggruppamento, in numerosi esempi e dimostrano le ottime prestazioni della fattorizzazione di affinità anche dove lo sfondo non è strutturato, cosa che non avviene utilizzando l'algoritmo del taglio normalizzato (figure 4.3(a)-4.3(b)). Shi e Malik hanno approcciato questo problema definendo misure di affinità più sofisticate, per le quali la classe di immagini non strutturate fosse minore.

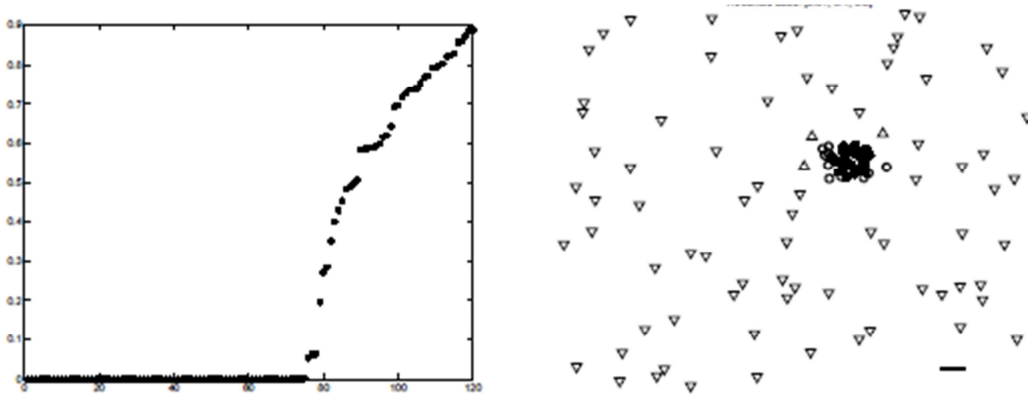


Figura 4.3(a):

Esempio di clustering attraverso l'algoritmo di fattorizzazione di affinità. A sinistra è mostrato l'andamento del vettore p , a destra i gruppi ottenuti attraverso l'algoritmo da una figura costituita da punti in primo piano e sullo sfondo. I triangoli capovolti indicano punti per i quali $p < 0.01$, per i triangoli $p \in (0.01, 0.1)$, i cerchi aperti indicano punti per cui $p \in (0.1, 0.5)$, per i cerchi chiusi infine $p > 0.1$. Da notare che i cerchi

sono indicativi dell'oggetto in primo piano, con i cerchi aperti che li circondano e i triangoli all'estrema periferia.

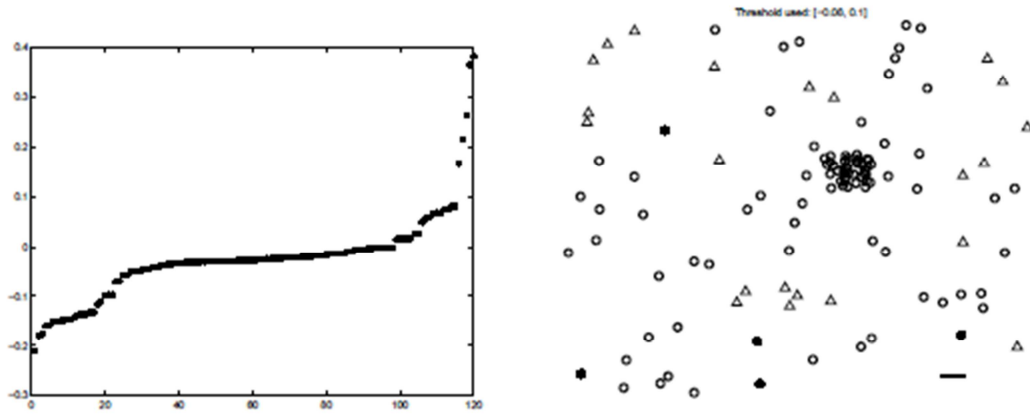


Figura 4.3(b):

Prestazioni dell'algoritmo del taglio normalizzato nello stesso insieme di dati di figura 4.3. Diversamente dal vettore p dell'algoritmo di fattorizzazione di affinità, l'autovettore x di A assume sia valori positivi che negativi. Il grouping è mostrato a destra, con i triangoli che indicano valori di $x < -0.08$, i cerchi aperti $x \in (-0.08, 0.1)$ e i cerchi chiusi $x > 0.1$. I simboli risultano mischiati e perciò non emergono gruppi separati.

La fenomenologia del grouping visivo descritta poco sopra è, di fatto, il risultato di un procedimento di riduzione di dimensionalità. Verranno ora illustrati i risultati ottenuti da questa analisi e gli esempi applicativi, che si sono rivelati utili per comprendere più nel dettaglio la possibilità di poter riprodurre, attraverso l'analisi spettrale di matrici di affinità calcolate utilizzando i modelli geometrici di Citti-Sarti, il fenomeno percettivo di grouping nello spazio $\mathbb{R}^2 \times S^1$.

4.3 Matrice di affinità

Nel precedente paragrafo sono state descritte le tecniche per estrarre informazioni di grouping visivo, in particolare lo spectral clustering. Queste prevedono l'esistenza di una matrice di affinità simmetrica e semi-definita positiva, di cui si andrà ad analizzare lo spettro.

Dati due elementi (i, j) della scena S , è possibile infatti assegnare un numero $A_{i,j}$ che indica quanto sono affini i due oggetti: quando sono simili, $A_{i,j}$ assume un valore alto, altrimenti $A_{i,j} \approx 0$. Questa funzione di affinità ha inoltre le seguenti proprietà:

$$\forall i, j \in S: \quad A_{i,j} \in [0,1], \quad A_{i,i} = 1, \quad A_{i,j} = A_{j,i}.$$

Le prime due proprietà impongono la normalizzazione di A , l'ultima rappresenta la simmetria della matrice.

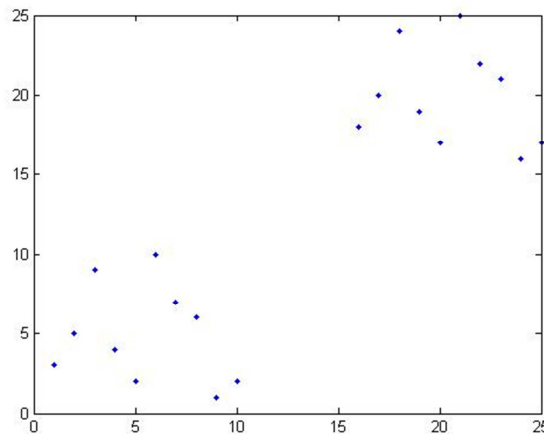


Figura 4.4: Insieme di punti nel piano

Come esempio, due punti in figura 4.4 sono ‘simili’ se vicini. È possibile quindi definire $A_{i,j}$ nel modo seguente:

$$(4.11) \quad A_{i,j} = e^{-d(i,j)^2}$$

$$d_{i,j}^2 = \frac{\|x_i - x_j\|^2}{d_0^2}$$

dove d_0 è la distanza di riferimento, sotto la quale due punti sono considerati simili e sopra la quale diversi, e x_i è il vettore di coordinate dei punti i .

Il fatto che segmenti tendono a raggrupparsi in strutture lineari quando allineati era stato notato negli anni '30 dagli psicologi della Gestalt, come mostrato nel capitolo 2. Un modo per definire l'affinità tra segmenti, seguendo sia le regole della Gestalt che le misure psico-fisiche degli anni '90 [10][12], assume che due segmenti lineari sono 'simili' quando sono vicini, allineati e co-circolari, cioè tangenti allo stesso cerchio. Una possibile implementazione della funzione di affinità è:

$$(4.12) \quad A_{i,j} = e^{-\frac{d_{i,j}^2}{d_0^2} - \frac{2 - \cos(2\alpha_i) - \cos(2\alpha_j)}{1 - \cos(2\theta_0)} - \frac{1 - \cos(2\alpha_i - 2\alpha_j)}{1 - \cos(2\delta\theta_0)}}$$

dove il primo termine nell'esponenziale rappresenta la dipendenza dell'affinità dalla distanza tra segmenti ($d_{i,j}$ in figura 4.5), il secondo termine penalizza la deviazione media dei segmenti lineari dall'essere allineati, e l'ultimo termine penalizza la non-cocircularità tra due segmenti. Le costanti $d_0, \theta_0, \delta\theta_0$ sono arbitrarie; d_0 assume buoni valori nell'intervallo che va dalla misura della spaziatura tra gli elementi in esame fino a cinque volte questo valore. Buoni valori per θ_0 vanno da $\pi/2$ a $\pi/10$, mentre $\delta\theta_0$ è tipicamente tra un quarto e la metà di θ_0 [10].

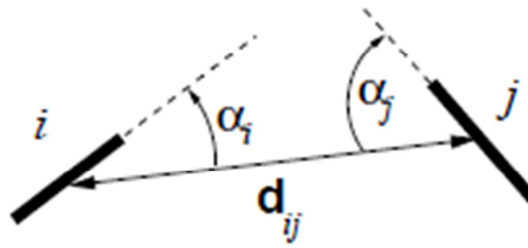


Figura 4.5:

Variabili coinvolte nel calcolo dell'affinità tra due segmenti i e j .

I metodi per simulare la fenomenologia percettiva del grouping prevedono l'esistenza della matrice di affinità, avente le caratteristiche sopra riportate. Verrà ora illustrata la costruzione di questa matrice: a partire dall'integrazione del sistema di equazioni differenziali (2.15), si ottiene un nucleo differenziale stocastico dipendente da alcuni parametri, quali numero di percorsi, numero di passi utilizzati in ognuno dei percorsi e dalla costante di diffusione σ^2 , oltre che dalle condizioni iniziali. Ad esempio, prendendo un nucleo con condizione di orientazione iniziale $\Theta_0 = 0$ si ricava un kernel con asse orizzontale.

Per il calcolo della matrice di affinità, invece di utilizzare la misura di distanza data dall'equazione (4.12), avendone a disposizione una che si basa su un modello corticale (quindi si assume che sia più rappresentativa), si utilizza questa distanza calcolata con i nuclei stessi. In particolare, per ogni punto si calcolano le distanze $dx, dy, d\theta$ da ogni altro punto del set di dati. Associando poi il nucleo calcolato in precedenza alla distanza tra i punti del dominio $\mathbb{R}^2 \times S^1$ e non alla funzione di similarità Gaussiana:

$w(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}$, che spesso viene utilizzata, si ricava l'affinità di quel punto con tutti gli altri, ottenendo proprio la Matrice di affinità.

L'esempio che segue descrive il concetto di affinità:

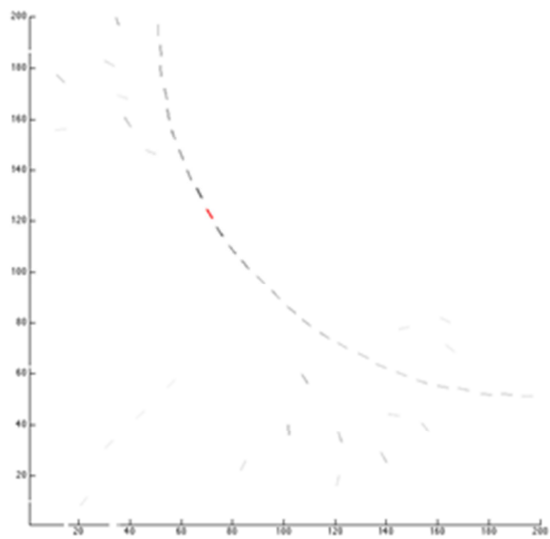
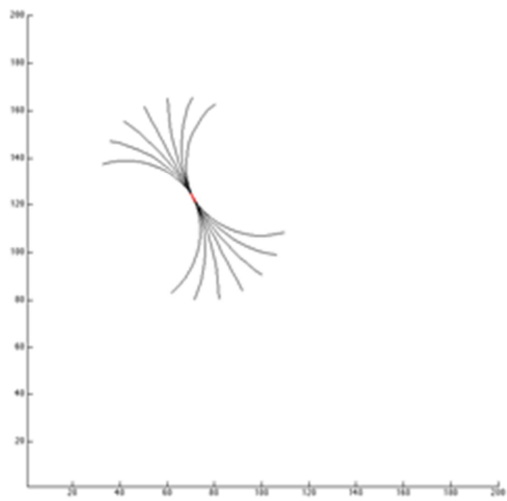
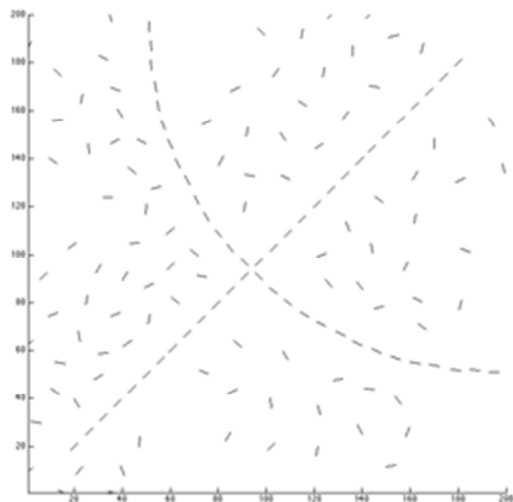


Figura 4.6

Sia dato lo stimolo (riportato in alto) caratterizzato da un'unità percettiva costituita da segmenti disposti lungo una retta, una seconda unità con segmenti lungo una curva e diversi segmenti random. Per ogni elemento, rappresentato in rosso come mostrato nella figura in basso, è possibile calcolare l'affinità con i segmenti vicini, attraverso un nucleo centrato proprio nell'elemento rosso (riportato in basso).

Nella serie di esempi sotto riportati, si calcola la Matrice di affinità e quindi l'affinità tra gli stimoli al variare dei parametri del nucleo stocastico. Si considera lo stimolo di figura:

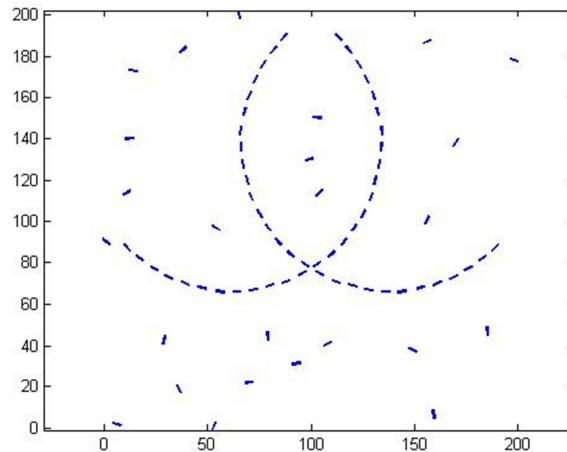


Figura 4.7:

Esempio di stimolo visivo composto da una serie di segmenti, alcuni disposti in maniera casuale, altri lungo due linee con diverse curvature. Quest'ultimi appartengono a due unità percettive, gli altri sono elementi di rumore all'interno del dominio $\mathbb{R}^2 \times S^1$.

nel quale ogni segmento è dotato di una certa orientazione e quindi può rappresentare l'attività neurale in risposta allo stimolo stesso. Per questo esempio, ho creato uno stimolo costituito sia da segmenti formanti due unità percettive distinte che disposti in posizione random. Calcolate le distanze $dx, dy, d\theta$ di ogni segmento con gli altri costituenti il set in

esame, si ricava la Matrice di affinità e l'affinità di ogni segmento con quelli adiacenti, come mostrato in figura 4.8.

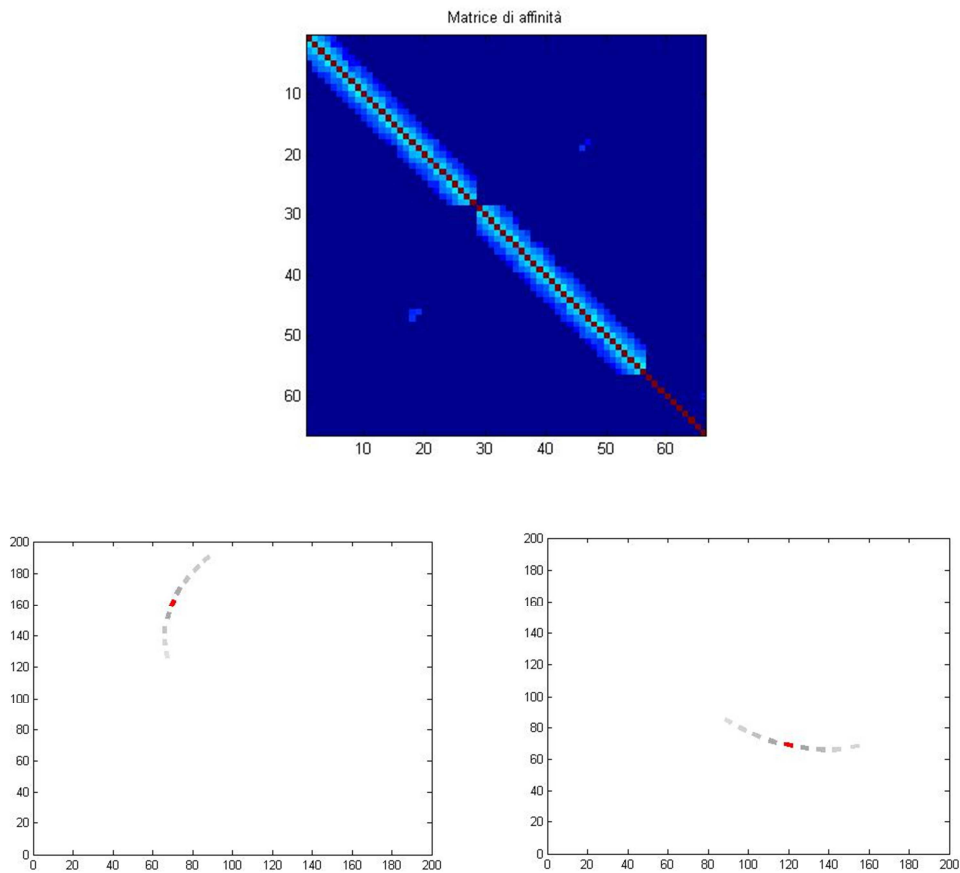


Figura 4.8:

A partire dallo stimolo di figura 4.7, si ricava la Matrice di affinità (riportata in alto) calcolata dal nucleo caratterizzato dai seguenti parametri: $\sigma^2=0.07$, 1000 percorsi e 40 step. Per ottenere la matrice si associa il nucleo alla distanza tra i punti del dominio. Come si nota dalla figura, emergono le due unità percettive, che andranno poi a costituire due gruppi distinti. In basso sono rappresentati due esempi di affinità tra il segmento in esame (in rosso) e quelli adiacenti.

L'esempio sopra riportato mostra la netta separazione dei dati nei due cluster principali, rappresentati dalle unità percettive corrispondenti. La Matrice di affinità è stata ottenuta a partire da un nucleo stocastico orizzontale caratterizzato da un elevato coefficiente di diffusione, adatto a

riconoscere segmenti disposti lungo una curva, e un elevato numero di step in modo tale che l'affinità non sia limitata ai soli segmenti vicini.

Si analizza ora la Matrice di affinità al variare di determinati parametri del nucleo, quali il numero di percorsi, il numero di passi utilizzati in ognuno dei percorsi e la costante di diffusione σ^2 , per vedere come cambiano le affinità tra stimoli adiacenti.

Il numero di step, o numero di passi in un percorso, rappresenta anche per la Matrice di affinità un indice di "lontananza": al diminuire di questa variabile la matrice non risulta definita oltre un certo raggio, come mostrato in figura 4.9.

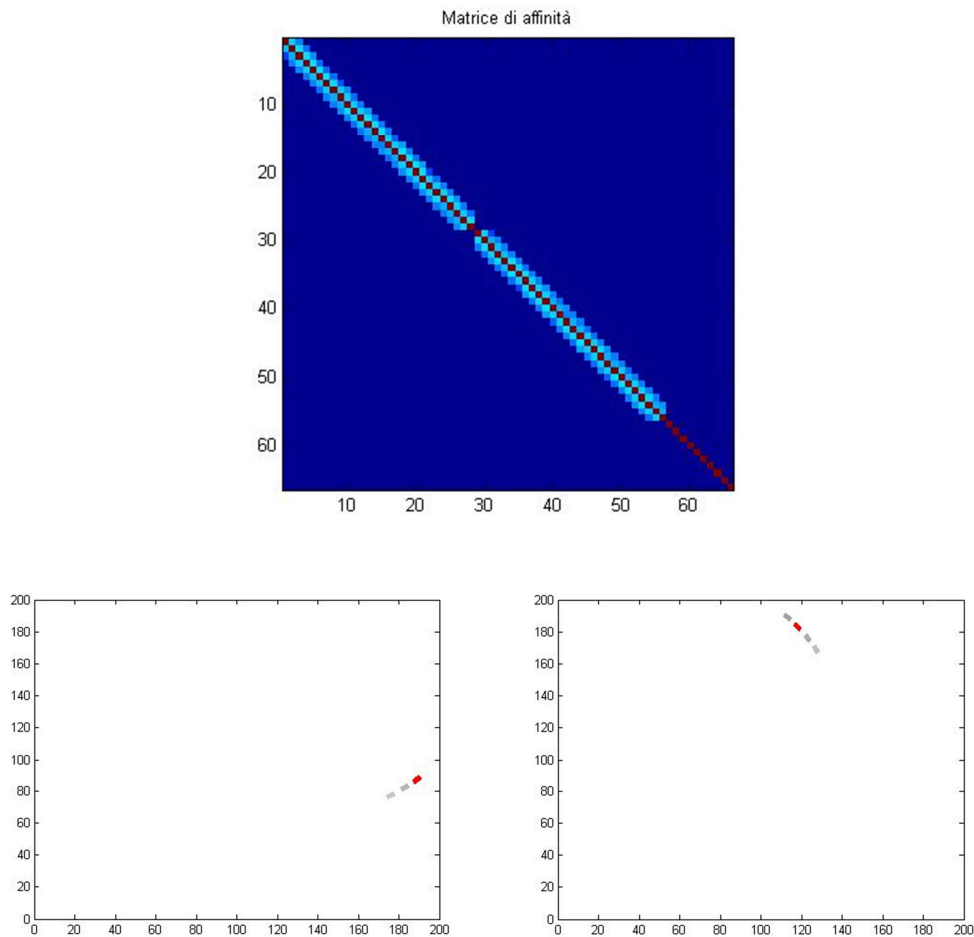


Figura 4.9:

In alto è riportata la matrice di Affinità, calcolata a partire da un nucleo caratterizzato da costante di diffusione $\sigma^2=0.07$, 1000 percorsi e 20 step. In essa si nota ancora la distinzione nei due gruppi, che risultano però meno ampi di prima e coinvolgono perciò un minor numero di punti. In basso due

esempi di affinità tra il segmento in esame (in rosso) e quelli adiacenti: queste immagini confermano che l'affinità risulta definita solo tra stimoli vicini a quello in esame.

Come riportato in figura, l'affinità è limitata e i segmenti oltre un certo raggio non vengono più coinvolti nel calcolo della Matrice.

Si considera ora l'omogeneità del kernel, definita principalmente dal numero di percorsi. Al diminuire di questi, il nucleo risulta discontinuo e l'affinità è sempre definita tra segmenti vicini; all'aumentare della variabile in esame il kernel e quindi la Matrice di affinità risultano più omogenei e definiti per un maggior numero di punti. Un numero di percorsi elevato è pertanto indicato per ottenere un buon nucleo e di conseguenza una Matrice di affinità più ordinata e uniforme.

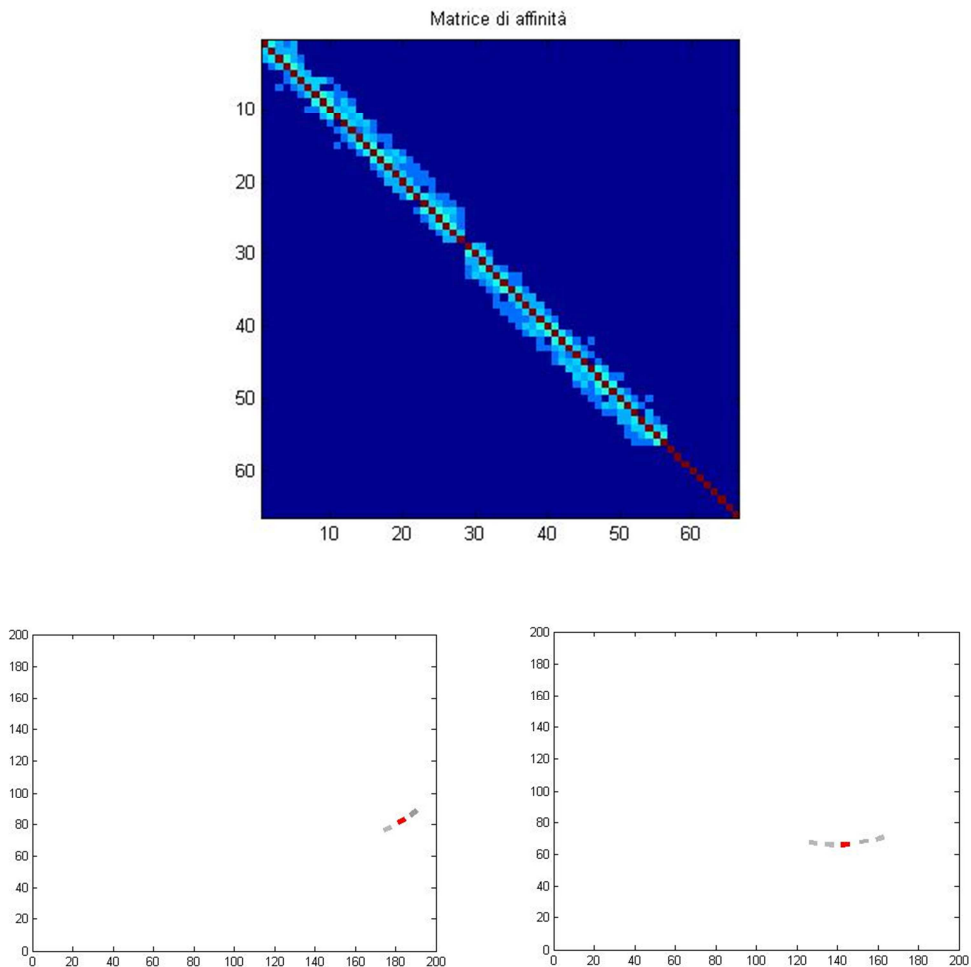


Figura 4.10:

In alto è riportata la Matrice di affinità, calcolata a partire da un nucleo caratterizzato da $\sigma^2=0.07$, 100 percorsi e 40 step. Questa matrice risulta variata rispetto a quella di partenza (figura 4.8), in cui i gruppi risultano meno omogenei e definiti.

In basso vi sono due esempi di calcolo di affinità di due segmenti con quelli adiacenti: l'affinità tra stimoli risulta definita in maniera più casuale essendo il nucleo di partenza più disordinato.

In questo caso, con un basso numero di percorsi, la Matrice di affinità non è più simmetrica: l'approssimazione del nucleo con un sistema di equazioni stocastiche (3.2) può portare a questo tipo di errore.

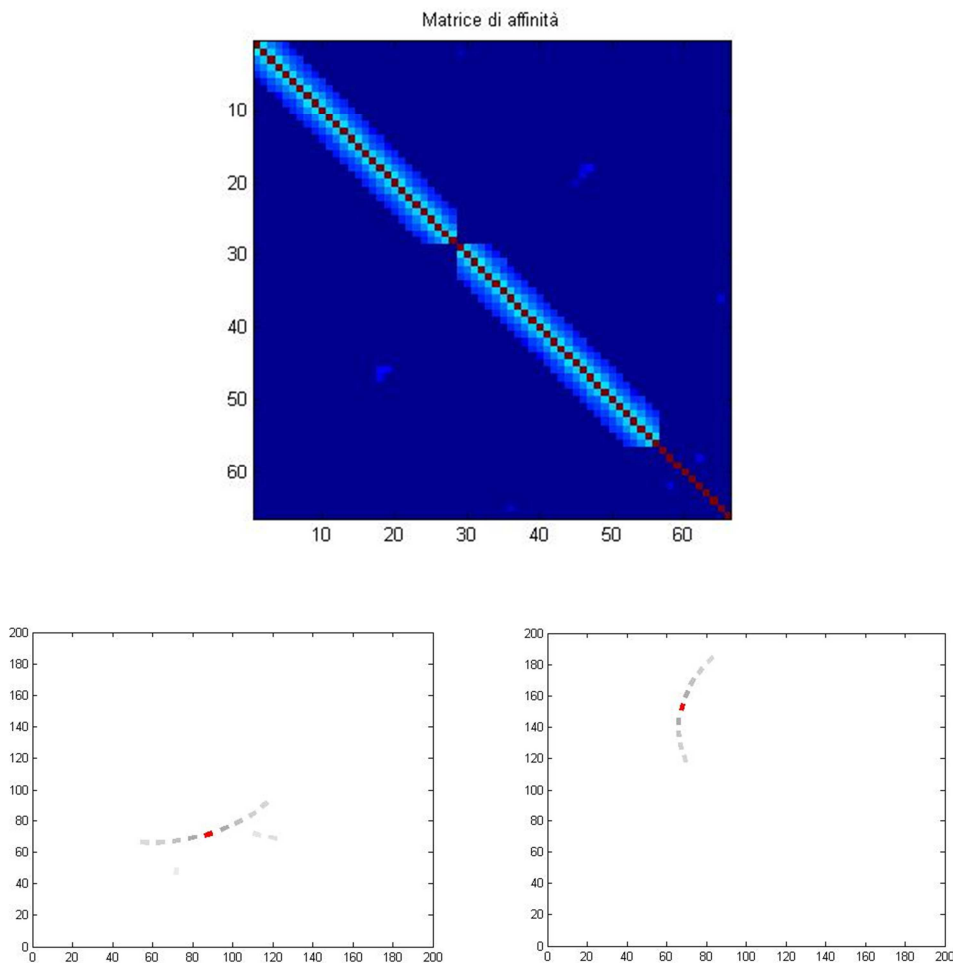


Figura 4.11:

La Matrice di affinità che si ottiene aumentando nel calcolo del nucleo il numero di percorsi da 100 a 10000 e mantenendo $\sigma^2=0.07$ e 40 step. Questa matrice risulta molto più omogenea e uniforme, rispetto a quella calcolata nel caso precedente (figura 4.10). In basso sono rappresentati due esempi di calcolo di affinità di due segmenti (in rosso) con quelli adiacenti: questa non è limitata ai soli elementi vicini.

Come mostrano le figure 4.10 e 4.11, aumentando il numero di percorsi aumenta l'affinità tra gli elementi dello stimolo.

Infine l'ultima variabile considerata è la costante di diffusione σ^2 : come mostrato in precedenza, al diminuire di questa il kernel degenera in una retta e l'affinità è limitata quasi esclusivamente a segmenti allineati, come mostrato in figura 4.12. All'aumentare di σ^2 , aumenta l'affinità tra segmenti disposti lungo percorsi curvilinei, come quello in esame.

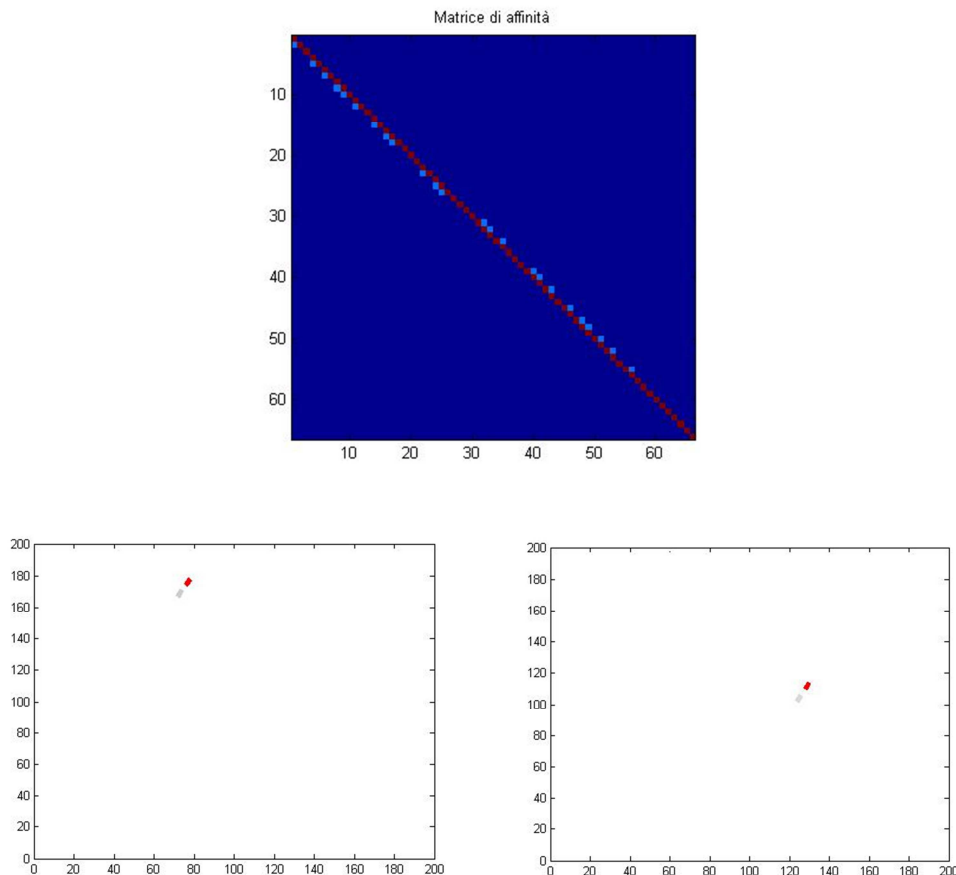


Figura 4.12:

Calcolo della Matrice di affinità a partire da un nucleo caratterizzato da $\sigma^2=0.01$, 1000 percorsi e 40 step. Essendo il coefficiente di diffusione basso, ogni elemento risulta affine con segmenti vicini e allineati, come mostrato anche negli esempi di calcolo di affinità di due segmenti con quelli adiacenti (in basso).

In questo esempio (figura 4.7), occorre un elevato coefficiente di diffusione per riconoscere segmenti disposti lungo una curva, un alto numero di percorsi per avere nucleo e matrice di affinità omogenei e un numero di step non troppo basso per evitare che l'affinità risulti definita solo tra stimoli vicini a quello in esame. In base allo stimolo, è quindi necessario adattare questi parametri, in modo da calcolare al meglio l'affinità tra gli stimoli e poter poi simulare la fenomenologia percettiva del grouping.

In alcuni casi, come ad esempio in presenza di distribuzioni puntiformi di punti molto densi, si può ottenere una matrice di affinità diagonale a blocchi, dove ogni blocco rappresenta una componente connessa del grafo (figura 4.14).

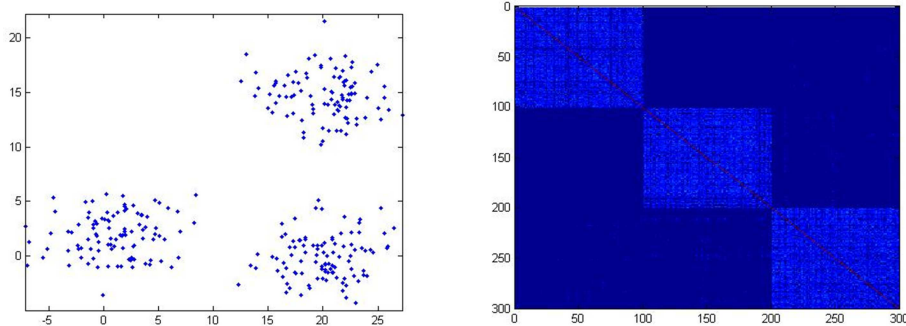


Figura 4.14:

Esempio di stimolo costituito da distribuzioni puntiformi e corrispondente matrice di affinità a blocchi, in cui si possono visualizzare i 3 gruppi in cui è possibile suddividere lo stimolo.

In questo caso la molteplicità dell'autovalore pari a zero è pari al numero di componenti connesse e i relativi autovettori sono dei vettori indicatori per ogni componente. In questo caso è semplice trovare dei cluster, considerando le componenti sconnesse del grafo. Ci sono poi casi in cui la matrice a blocchi è fortemente perturbata (figura 4.15), il grafo è connesso ma in realtà sono distinguibili cluster che sono tra loro praticamente sconnessi, a meno di qualche connessione. In questo caso un solo autovalore sarà uguale a zero, ma altri saranno piccoli e se la perturbazione non è troppo grande si è in grado di trovare facilmente dei cluster [29].

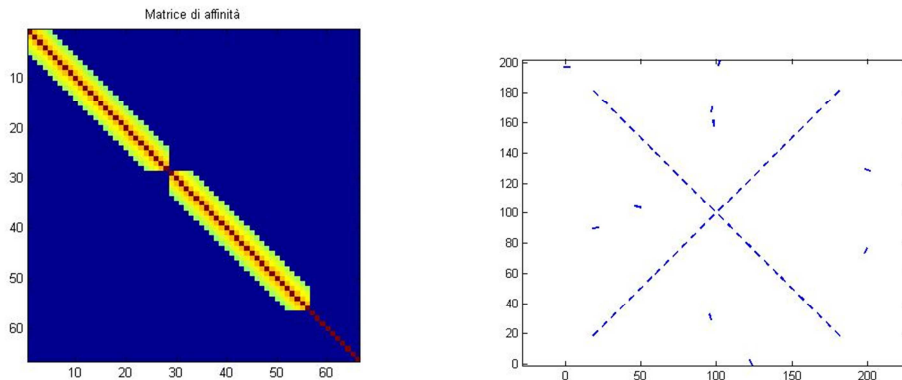


Figura 4.15:

Esempio di matrice di affinità a blocchi perturbata (a sinistra), calcolata a partire da un nucleo caratterizzato dai seguenti parametri: $\sigma^2=0.01$, 1000 percorsi e 40 step.. Questa si ottiene considerando uno stimolo composto da segmenti disposti lungo due rette, oltre a alcuni segmenti di rumore (rappresentato a destra). Si distinguono nettamente i due cluster, completamente sconnessi, che rappresentano i segmenti che appartengono alle due unità percettive.

Verrà ora illustrata una serie di esperimenti per il testing dello Spectral clustering, dove si verifica anche sotto quali condizioni (valore dei parametri) si ottiene una buona qualità dei risultati.

4.4 Esempi applicativi

In questa sezione vengono presentati alcuni esempi di estrazione di informazioni di grouping visivo, attraverso le tecniche presentate di Spectral clustering. Attraverso delle simulazioni si creano dei training set; l'algoritmo non conosce queste distribuzioni e deve ricavarle proprio dall'insieme di dati. Si analizzano casi semplici ma che sintetizzano problematiche che si possono incontrare nei casi pratici. In ogni simulazione si cerca di partizionare i dati nei cluster corretti, analizzando i parametri per ottenere questi raggruppamenti. Vengono analizzati due classi di problemi: l'articolazione figura/sfondo e l'estrazione di oggetti con tecniche di riduzione di dimensionalità.

Nella prima, si considera l'esempio di un disco nero su sfondo bianco (rappresentato da due cerchi concentrici e con raggi che differiscono di poco) per valutare la separazione del background dal foreground. In questo caso, si utilizza l'algoritmo di Perona-Freeman: a partire dalla matrice di affinità si calcola l'autovettore p corrispondente al suo più grande autovalore e si divide lo sfondo dall'oggetto proprio in base ai valori assunti da p .

Nella seconda, si creano stimoli composti da elementi di rumore e da segmenti che assumono particolari geometrie (si passa da disposizioni rettilinee a curvilinee). In questa, dopo aver calcolato il nucleo e la matrice di affinità corrispondente, si calcola la matrice di transizione P (da cui si ricava il Laplaciano normalizzato L_{rw} ; equazione(10)) e i corrispondenti autovalori e autovettori significativi per poter fare clustering.

4.4.1 Articolazione figura-sfondo

Nel secondo capitolo sono state presentate le teorie della Gestalt e le sue regole, secondo le quali stimoli visivi distinti tra loro possono essere percepiti come un'unica unità percettiva. In riferimento alle percezioni

visive, tali principi hanno il fine di spiegare quali parti dello stimolo sono associati a formare qualità percettive salienti (bordi, figure, sfondi, qualità cromatiche e acromatiche). L'articolazione figura-sfondo è uno dei fondamentali processi che stanno alla base di ogni esperienza visiva, e rappresenta la tendenza del campo percettivo a dividersi in ciò che è primario e importante da ciò che è secondario. Per figura si intende quegli insiemi di elementi che sono delimitati da un contorno (reale o apparente) che all'interno dell'immagine attira la nostra attenzione; per sfondo si intende tutto ciò che appare lontano e a distanza indefinita, e fa da "cornice" alla figura [18]. Le leggi di articolazione figura-sfondo descrivono quali parti del campo percettivo diventano figura e quali lo sfondo. Alcune di queste sono:

- 1) la legge dell'inclusione: a parità di altre condizioni, viene percepita come figura la regione di spazio chiusa o circondata da altre aree;
- 2) la legge dell'area relativa: la zona più piccola tende ad emergere come figura, la più grande come sfondo;
- 3) la legge dell'orientamento: le zone in cui gli assi sono allineati con gli assi verticali e orizzontali diventano regione.

Quando nessuna delle condizioni riportate sopra privilegia una parte del campo delle altre, si ha una situazione di ambiguità, nella quale domina l'instabilità e la continua reversibilità del rapporto figura-sfondo [18]. Un esempio è rappresentato dal vaso di Rubin (figura 4.16), dove risulta impossibile percepire contemporaneamente i due volti e il candelabro: quando prevale il risalto dei due bordi, ad esempio, il candelabro perde il suo carattere di figura.

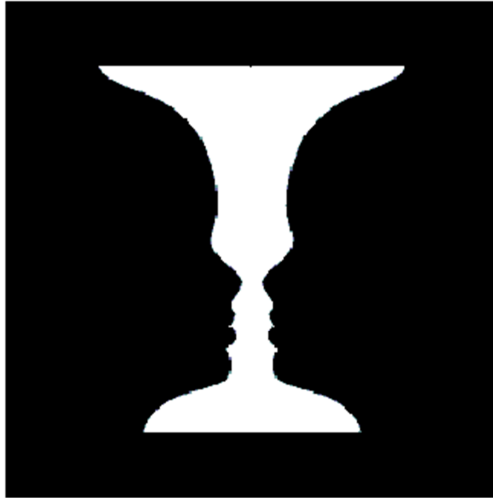


Figura 4.16: Esempio di figura ambigua: il vaso di Rubin.

Per analizzare questo problema di articolazione di foreground-background ho considerato l'esempio semplice di un disco nero su sfondo bianco (figura 4.17).

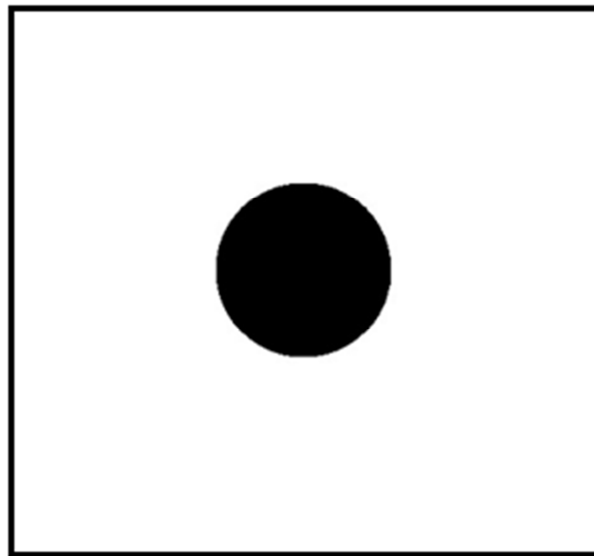


Figura 4.17:

Esempio di articolazione figura-sfondo: disco nero su sfondo bianco.

Convolvendo questa immagine con la parte reale di un filtro di Gabor (equazione 1.4) per diverse orientazioni e sommando i risultati in modo da ottenere un'informazione completa, ho osservato che l'immagine può essere

scomposta in due cerchi, uno più interno e uno più esterno, che permettono di separare ciò che è figura da ciò che è sfondo (figura 1.8).

Assumendo che le cellule siano connesse come da schema di connettività spazio-orientazione precedentemente descritto, si può notare come la risposta cellulare vari con l'orientazione: in figura 1.8 vengono rappresentate le zone di massima risposta per $\Theta = 0^\circ, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3}{4}\pi$.

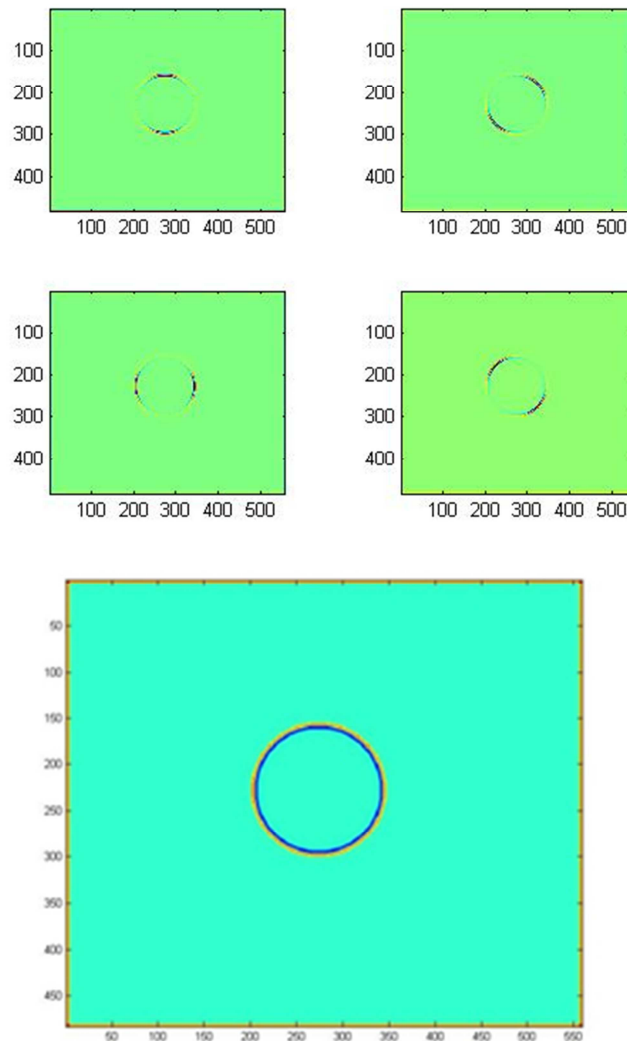


Figura 4.18:

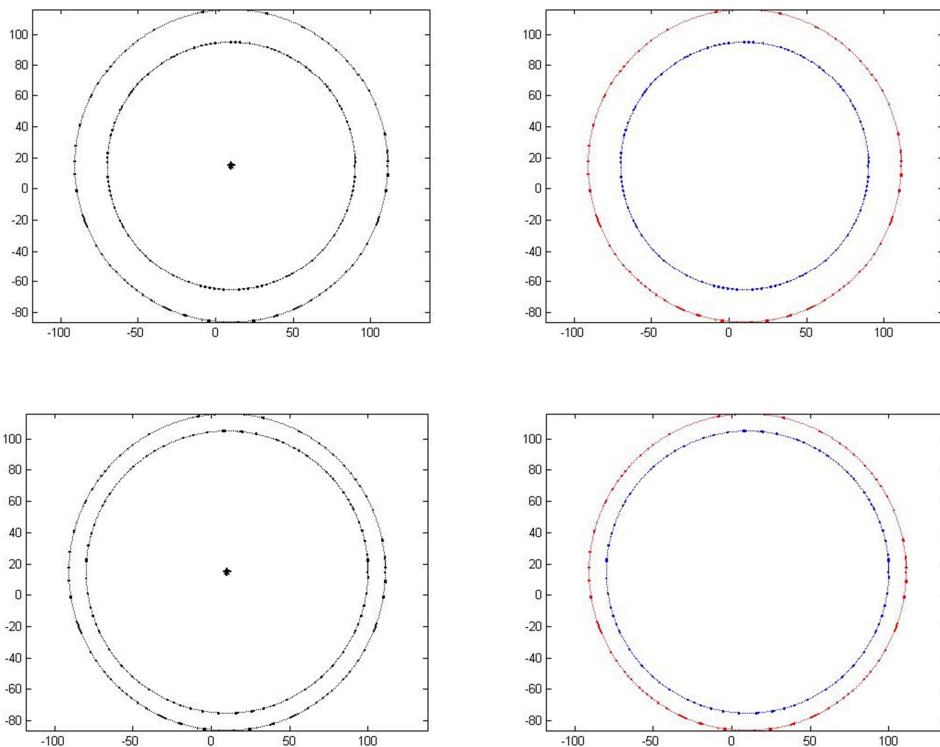
Convolutione dell'immagine con un filtro di Gabor pari, al variare di Θ .

In basso, l'unione dei risultati precedenti.

Ad esempio, per un angolo Θ pari a 0° , la cellula risponde massimamente nel bordo superiore e inferiore del cerchio, essendo l'asse di simmetria cellulare orizzontale e quindi tangente al bordo in quella posizione.

Rappresentando quindi le risposte di cellule semplici attraverso due serie di segmenti circolari (cerchi concentrici) e la connettività delle cellule con il concetto di affinità, si può effettuare la segmentazione figura-sfondo, dove l'autovalore più significativo è sempre la figura, il resto rappresenta lo sfondo.

Lo stimolo è composto da segmenti disposti lungo due cerchi concentrici, ai quali ho applicato il III° algoritmo, quello di Perona-Freeman: a partire dalla matrice di affinità ho calcolato l'autovettore p corrispondente al suo più grande autovalore. È possibile così suddividere le due parti dell'immagine proprio in base ai valori assunti da p : valori pari a zero fanno parte dello sfondo, quelli maggiori di zero dell'oggetto in esame.



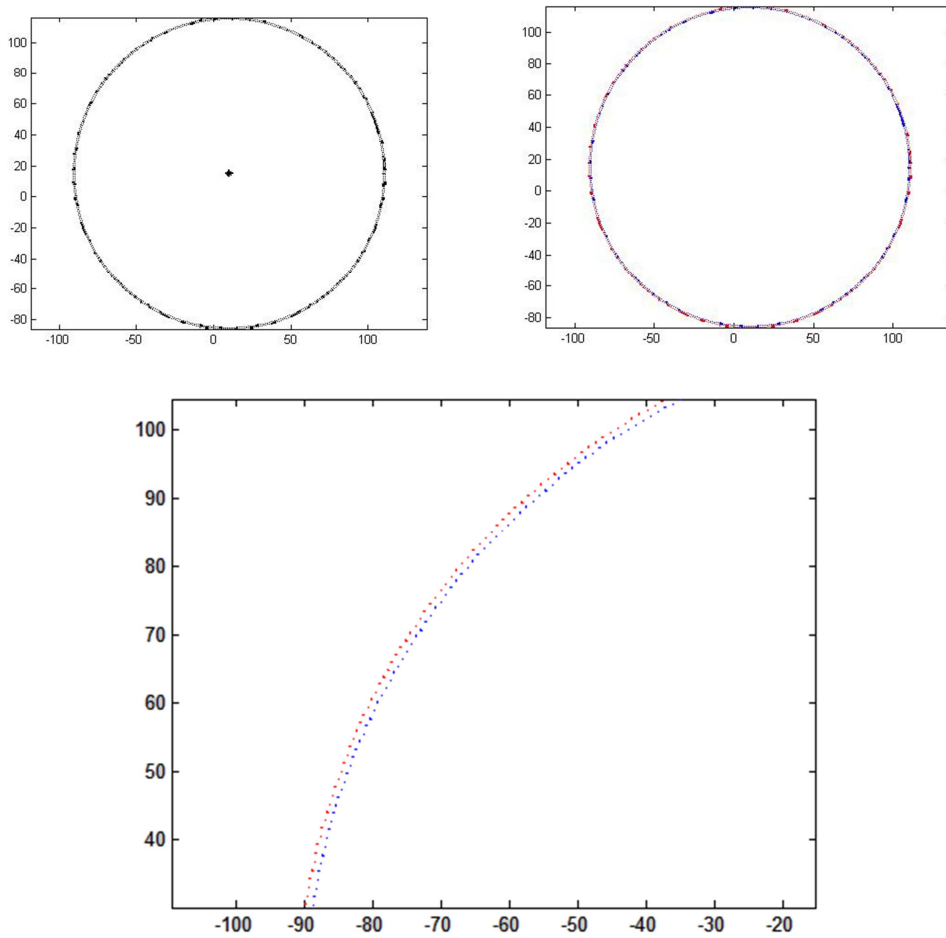


Figura 4.19:

Esempi di segmentazione dei due cerchi all'aumentare del raggio del cerchio interno. In tutti e tre i casi, i valori di p pari a zero sono rappresentati in rosso e fanno parte dello sfondo, quelli maggiori di zero (in blu) l'oggetto in esame. Nell'ultima immagine in basso è riportato uno zoom della figura precedente, per mostrare il corretto risultato e la separazione oggetto-sfondo, nonostante i cerchi siano praticamente coincidenti. Il nucleo è stato calcolato con $\sigma^2=0.006$, 1000 percorsi e 15 step.

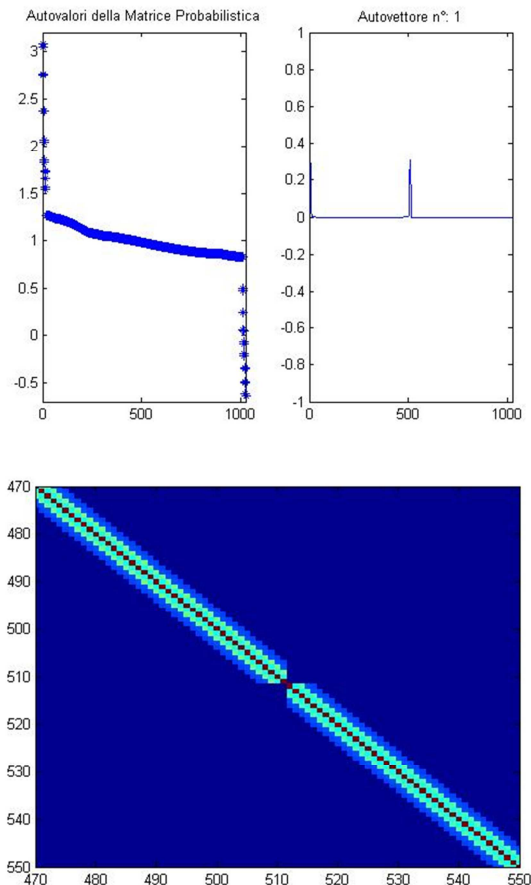


Figura 4.20:

In alto, è riportato lo spettro degli autovalori e del primo autovettore p , cioè quello associato all'autovalore più significativo. Si definisce quindi 'foreground' F l'insieme di oggetti i i cui corrispondenti p_i sono diversi da zero (o meglio $p > \varepsilon > 0$). In basso, zoom della matrice di affinità, nella zona di distinzione tra i due gruppi: analizzando i valori assunti dalla matrice di affinità, questi sono più elevati nel primo blocco, che corrisponde all'oggetto, mentre il secondo rappresenta lo sfondo.

Come si nota dalle figure, l'algoritmo riesce a separare correttamente i due cerchi, in tutti i casi esaminati. Analizzando i valori di p ho così verificato che gli autovettori più salienti sono proprio quelli in corrispondenza del cerchio interno, in cui i punti sono più vicini tra loro.

Un altro stimolo considerato è quello di figura 4.21: un'immagine costituita dalla sovrapposizione di due cerchi bianchi su sfondo nero.

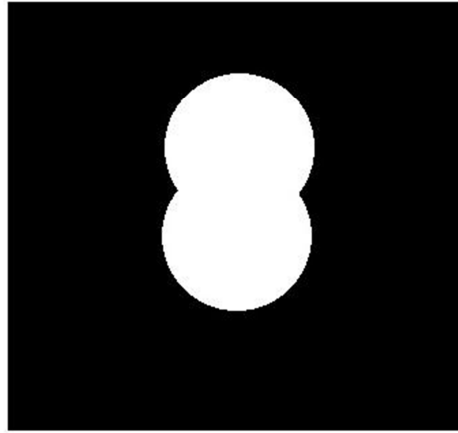
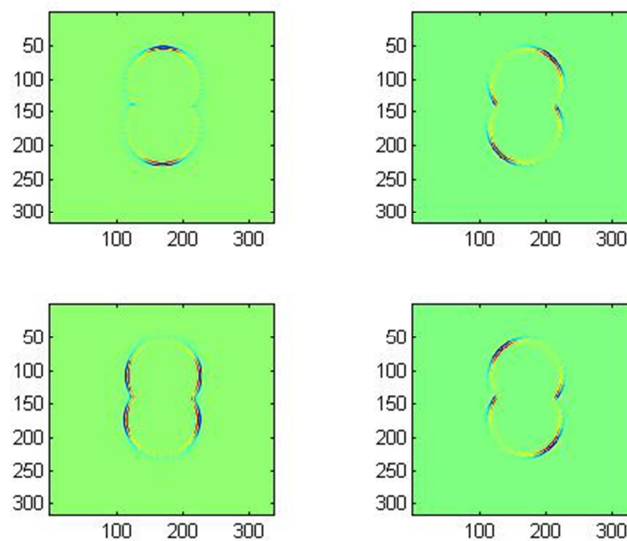


Figura 4.21:

Esempio di articolazione figura-sfondo: forma con angoli, costituita dalla sovrapposizione di due cerchi.

Agendo come visto precedentemente, si convolve questa immagine con un filtro di Gabor pari (equazione 1.4) per diverse orientazioni, ottenendo due forme, una più interna e una più esterna.

In figura 4.22 vengono rappresentate le zone di massima risposta per una determinata orientazione ($\Theta = 0^\circ, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$).



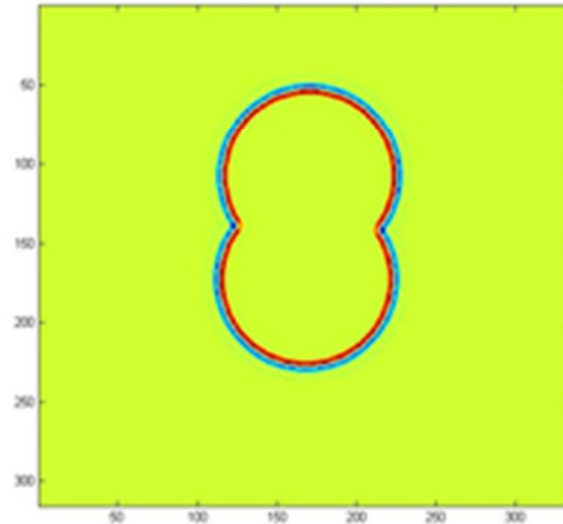


Figura 4.22:

Convoluzione dell'immagine con la parte reale di un filtro di Gabor, al variare di Θ . In basso, l'unione dei risultati precedenti.

Rappresentando quindi le risposte di cellule semplici attraverso queste due serie di segmenti e la connettività cellulare con il concetto di affinità, è possibile anche in questo caso effettuare la segmentazione figura-sfondo.

Utilizzando l'algoritmo di Perona-Freeman, ho calcolato l'autovettore p , corrispondente al suo più grande autovalore, e in base a questo ho segmentato l'immagine.

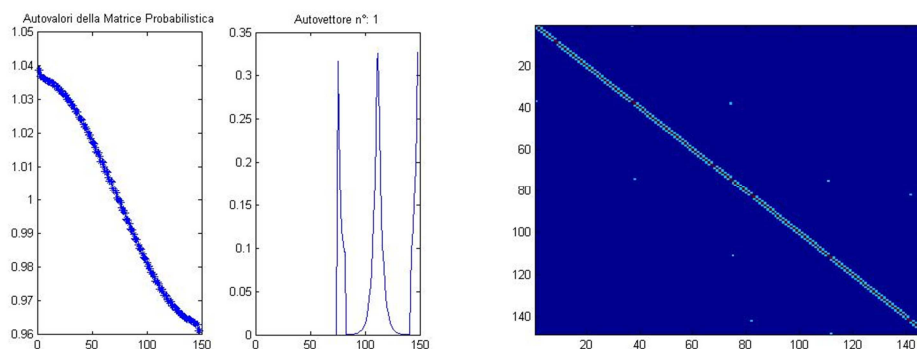


Figura 4.23:

A sinistra, grafico dello spettro degli autovalori e del primo autovettore p , cioè quello associato all'autovalore più significativo. Si definisce quindi

‘foreground’ F l’insieme di oggetti i i cui corrispondenti p_i sono diversi da zero (o meglio $p > \varepsilon > 0$). A destra, la matrice di affinità.

Nella Matrice di affinità è possibile identificare quattro gruppi: utilizzando infatti il nucleo del modello di connettività, questo non è in grado di rendere affini segmenti disposti lungo un percorso discontinuo, come a livello dell’asse di simmetria orizzontale immaginario che divide a metà l’immagine. Il nucleo è così in grado di percepire la forma esterna con continuità perché associata, in base all’algoritmo utilizzato, allo sfondo; mentre per la forma interna riconosce solo un semicerchio, fermandosi proprio a livello della discontinuità (figura 4.24). Assumendo però che in altre zone della corteccia esistano cellule che rispondono ad angoli o combinazioni di orientazioni [21] è possibile simularne il comportamento, modificando artificialmente la Matrice di affinità, ottenendo così la corretta articolazione figura-sfondo.

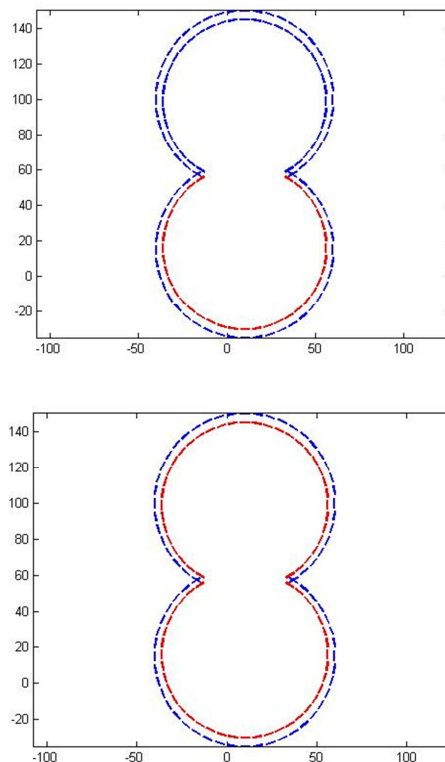


Figura 4.24:

Considerando come Matrice di affinità i nuclei del modello di connettività spazio-orientazione sopra riportato, questa non è in grado di percepire come continui stimoli contenenti variazioni in angolo, come si può vedere dall'immagine a sinistra: considerando la forma interna viene percepito solo uno dei due semicerchi. Attraverso l'assunzione dell'esistenza di cellule in grado di rispondere a discontinuità, ho modificato artificialmente l'affinità e verificato la corretta separazione delle due forme (immagine a destra).

In entrambi i casi analizzati, l'algoritmo riesce a separare correttamente la forma interna (l'oggetto) da quella esterna (lo sfondo) e gli autovettori più salienti sono risultati proprio quelli in corrispondenza del cerchio interno.

4.4.2 Estrazione di figure attraverso metodi di riduzione di dimensionalità

Per questi esempi ho creato diversi tipi di stimoli, in ognuno dei quali sono presenti due unità percettive costituite da segmenti, aventi curvature variabili. All'interno di questi stimoli vi sono inoltre dei segmenti di rumore, rappresentati da elementi in posizione casuale all'interno del dominio $\mathbb{R}^2 \times S^1$.

Come per gli esempi utilizzati per il calcolo della Matrice di affinità al variare dei parametri, ogni segmento è orientato e rappresenta la risposta all'attività neurale in risposta allo stimolo stesso.

Nei casi analizzati, si passa da una distribuzione dei dati rettilinea a una configurazione curvilinea, e la si partiziona attraverso il II° algoritmo di Spectral clustering modificato, cercando la soluzione ottimale al variare di determinati parametri. Infatti, per raggruppare i punti in cluster non si utilizza l'algoritmo del K-means (punto 6) dell'algoritmo) ma un'euristica diversa, avendo il K-means lo svantaggio di dover scegliere il numero di

partizioni k . A partire dall'algoritmo del Normalizer Spectral clustering, lo si modifica per fare in modo di calcolare automaticamente il numero ottimale di gruppi.

IV° algoritmo: Normalized Spectral clustering Modificato

Input: matrice di similarità S

- 1) Costruire un grafo di similarità completamente connesso. Sia W la Matrice di affinità, calcolata associando il nucleo alla distanza tra i punti del dominio $\mathbb{R}^2 \times S^1$.
- 2) Normalizzare la Matrice di affinità per righe, dividendone ogni riga per la sua somma (in questo modo si ottiene un'interpretazione probabilistica della Matrice).
- 3) Calcolarne autovalori e autovettori u_1, \dots, u_n .
- 4) Prendere i primi k autovalori significativi e i corrispondenti autovettori.
- 5) Sia $U \in \mathbb{R}^{n \times k}$ la matrice contenente u_1, \dots, u_k come colonne.
- 6) Calcolare i massimi valori che la matrice U assume per riga: la posizione $i = 1, \dots, k$ del massimo indica il cluster C_1, \dots, C_k di appartenenza dell'elemento in esame. Cluster costituiti da un unico elemento si rappresentano come appartenenti ad un unico gruppo.

Output: Cluster C_1, \dots, C_k .

Questo algoritmo supera il limite del K-means, calcolando automaticamente il numero k di cluster.

Per fare i raggruppamenti si utilizzano quindi autovalori e i corrispondenti autovettori significativi, calcolati dalla Matrice di affinità normalizzata.

A partire dall'esempio di figura 4.25, si sono calcolati: kernel, matrice di affinità, autovalori e autovettori significativi della Matrice di affinità normalizzata (figura 4.26) per ottenere i raggruppamenti corretti al variare dei parametri, in particolare della costante di diffusione e del numero di step.

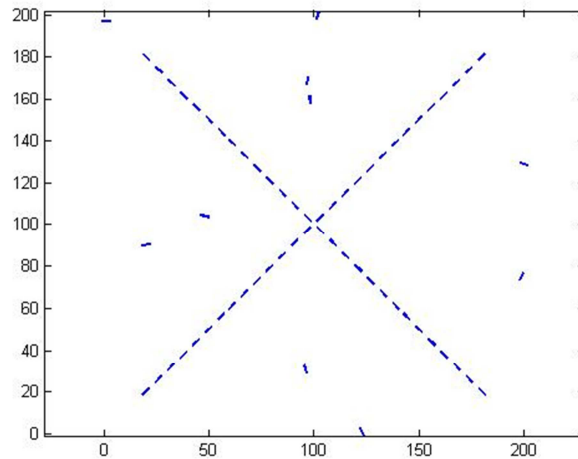


Figura 4.25:

Esempio di stimolo visivo, costituito da due unità percettive che sono formate da segmenti orientati disposti lungo due rette e 10 segmenti random nel dominio $\mathbb{R}^2 \times S^1$.

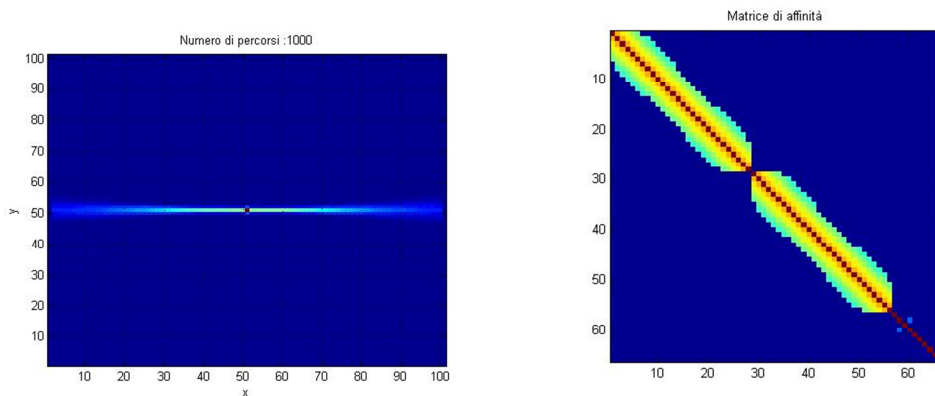


Figura 4.26:

A sinistra è riportato il nucleo stocastico orizzontale, caratterizzato da $\sigma^2=0.01$, 1000 percorsi e 50 step. A destra la Matrice di affinità, in cui si può distinguere la presenza di due cluster separati.

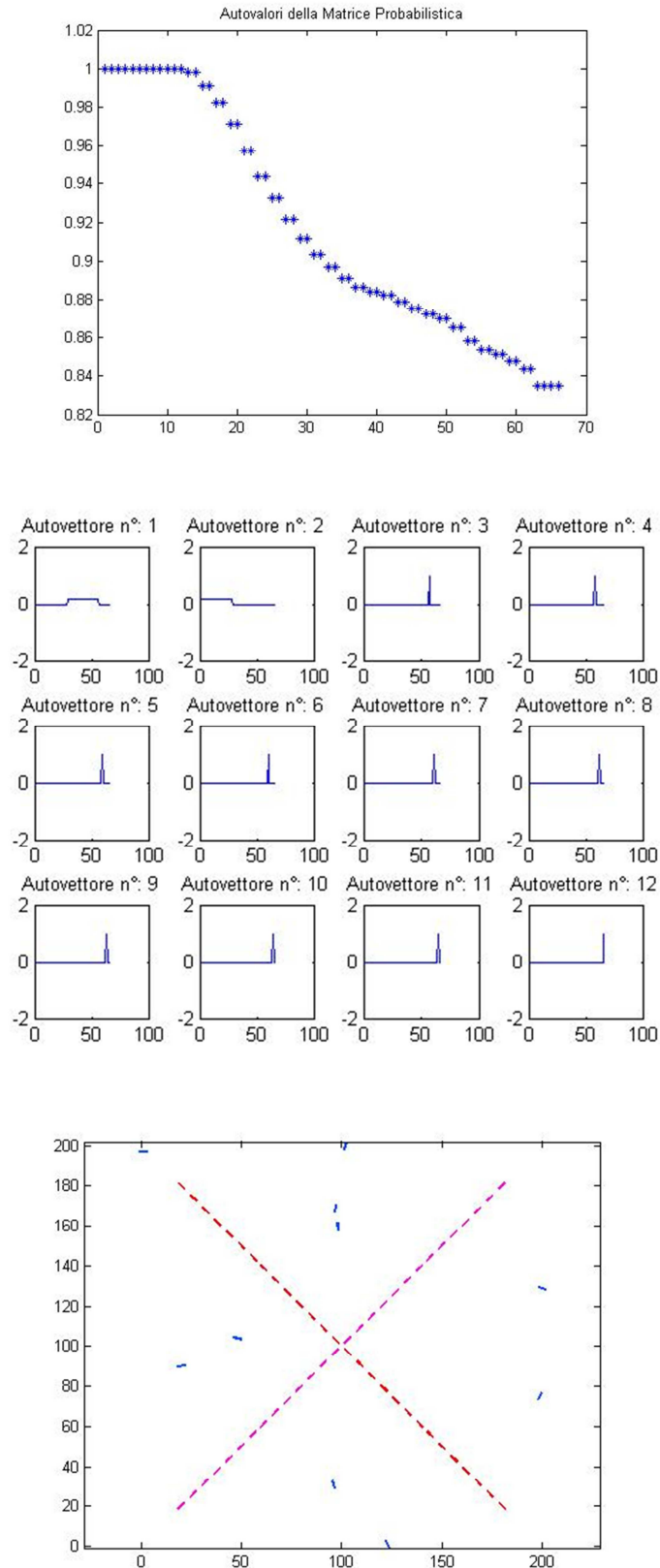


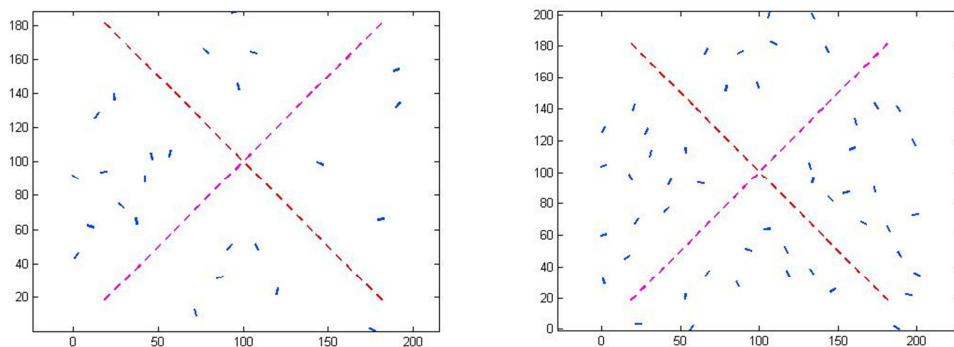
Figura 4.27

A partire dall'alto, il grafico degli autovalori e degli autovettori significativi, dove si nota la presenza dei due macro-gruppi (rappresentativi

delle due unità percettive) e di 10 gruppi costituiti da un unico elemento, uniti nell'algoritmo a formare il terzo cluster. Infine, il risultato del clustering: in rosso e viola sono rappresentate le due unità percettive, in blu i segmenti di rumore.

Lo spettro riportato in alto in figura 4.27 è ricco di informazioni: il primo e l'undicesimo autovettore sono costanti e rappresentano i due gruppi che individuano i diversi segmenti disposti lungo le rette; inoltre sono presenti altri dieci autovettori con un unico elemento diverso da zero: ognuno di questi rappresenta i segmenti di rumore, che vengono tutti raggruppati nel cluster blu.

Come si può notare dalle figura sopra riportate, l'algoritmo riesce a partizionare bene i dati in tre cluster distinti, unendo in un unico gruppo l'insieme dei segmenti di rumore. All'aumentare del rumore, mantenendo i valori delle variabili, l'algoritmo si comporta correttamente, riuscendo a separare i cluster tranne in caso di molto rumore, in cui qualche segmento viene erroneamente associato a uno dei due gruppi principali (figura 4.28).



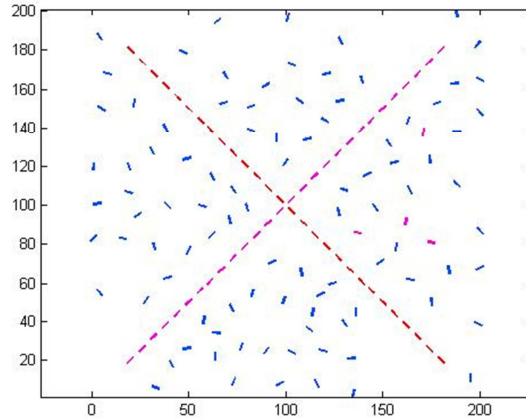
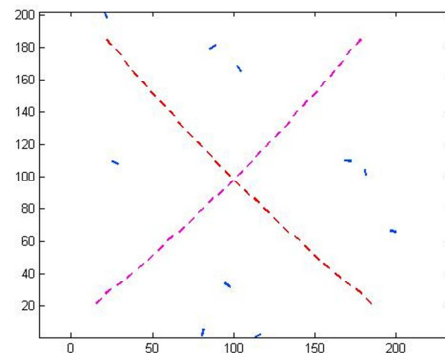
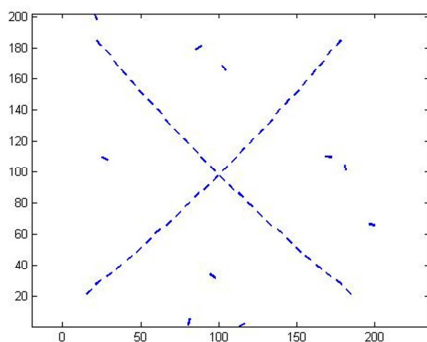


Figura 4.28:

Risultato del clustering all'aumentare del numero di segmenti di rumore (10, 25, 50 rispettivamente) nel dominio, rappresentati in blu. In rosso e viola sono rappresentate le due unità percettive. In caso di stimolo molto rumoroso, qualche segmento viene percepito in modo errato come appartenente a uno degli altri due cluster.

Uno dei parametri di interesse in questa analisi è la curvatura dello stimolo, definita come l'inverso del raggio di curvatura. Questa è direttamente proporzionale a $\frac{1}{2k}$, dove k è costante: quando assume valore 0 la curvatura risulta massima, per k elevati quest'ultima tende a 0 e i segmenti risultano disposti lungo una retta, come nello stimolo di figura 4.25.

Altri esempi sono stati calcolati variando proprio la curvatura delle due unità percettive, ottenendo i risultati riportati in seguito.



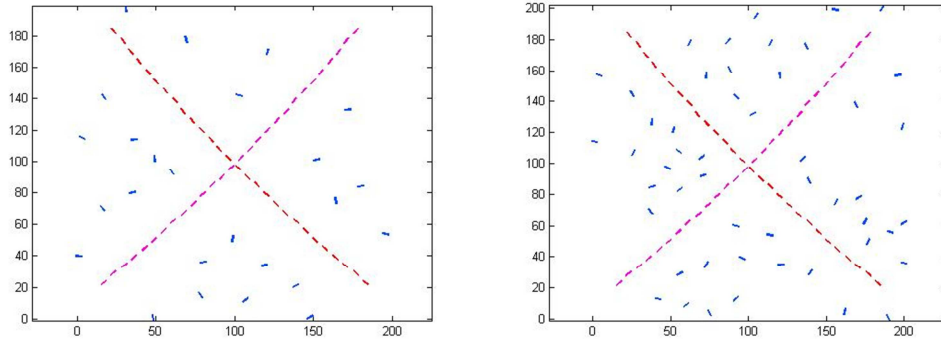


Figura 4.29:

In alto a sinistra è riportato un esempio di stimolo le cui unità percettive hanno curvatura pari a 0.0008 e in seguito sono mostrati i risultati del clustering all'aumentare del numero di segmenti di rumore (10, 25, 50 rispettivamente) nel dominio, rappresentati in blu. In rosso e viola sono rappresentate le due unità percettive. In tabella I sono illustrati i valori dei parametri usati per il calcolo del nucleo e della Matrice di affinità, in modo tale da ottenere un buon clustering.

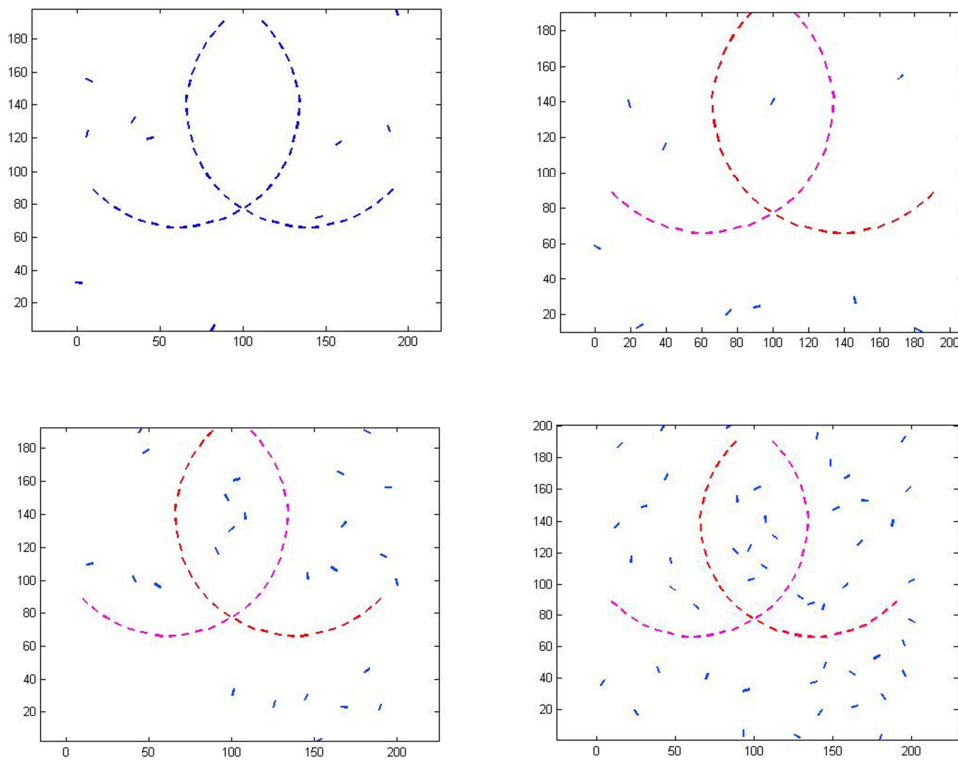


Figura 4.30:

In alto a sinistra è riportato un esempio di stimolo le cui unità percettive hanno curvatura pari a 0.0127 e in seguito sono mostrati i risultati del clustering all'aumentare del numero di segmenti di rumore (10, 25, 50 rispettivamente) nel dominio, rappresentati in blu. In rosso e viola sono rappresentate le due unità percettive. In tabella I sono illustrati i valori dei parametri usati per il calcolo del nucleo e della Matrice di affinità, in modo tale da ottenere un buon clustering.

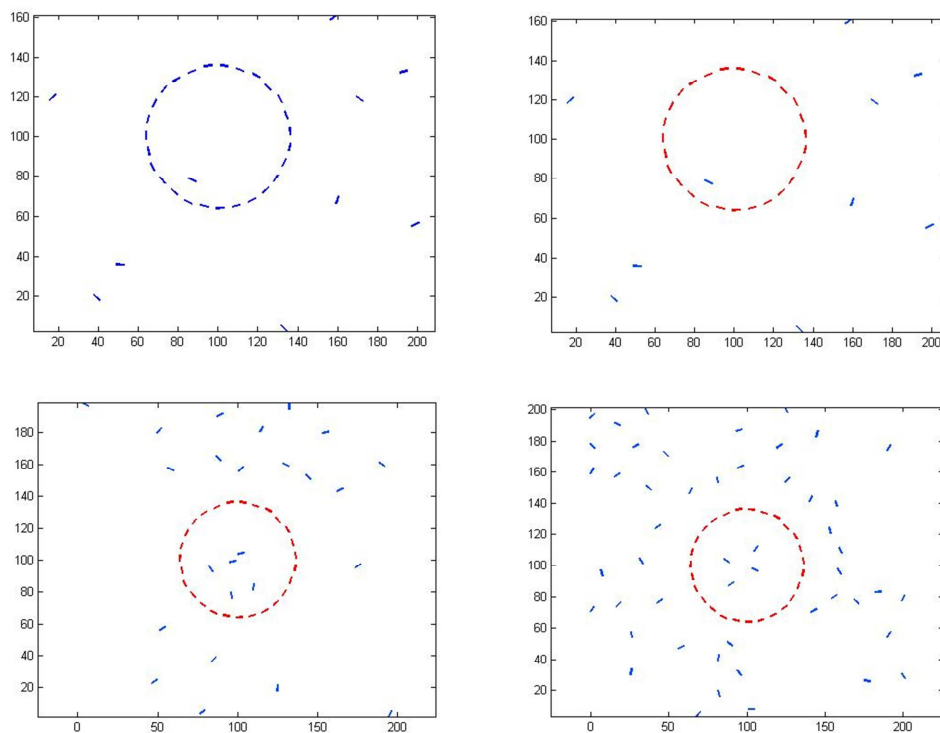


Figura 4.31:

In alto a sinistra è riportato un esempio di stimolo le cui unità percettive hanno curvatura pari a 0.0218 (configurazione a cerchio) e in seguito sono mostrati i risultati del clustering all'aumentare del numero di segmenti di rumore (10, 25, 50 rispettivamente) nel dominio, rappresentati in blu. In rosso e viola sono rappresentate le due unità percettive. In tabella I sono illustrati i valori dei parametri usati per il calcolo del nucleo e della Matrice di affinità, in modo tale da ottenere un buon clustering.

In tabella I si riassume l'analisi parametrica effettuata sugli stimoli:

Tipo di stimolo: curvatura	Rumore (n° segmenti)	Coefficiente di diffusione σ^2	Numero di step
0	10	0.01	50
0	25	0.01	50
0	50	0.01	50
0.0008	10	0.01	50
0.0008	25	0.01	50
0.0008	50	0.01	50
0.0127	10	0.07	35
0.0127	25	0.07	35
0.0127	50	0.075	35
0.0218	10	0.09	30
0.0218	25	0.09	30
0.0218	50	0.09	30

Come si nota dai valori assunti dai parametri, per uno stimolo rettilineo (figura 4.25) o con curvatura molto bassa (figura 4.29), si ottiene una buona segmentazione utilizzando un kernel avente basso coefficiente di diffusione, che così risulta praticamente rettilineo (figura 4.26). Inoltre il numero di step, che definisce il raggio del nucleo, è elevato per poter riconoscere segmenti allineati come appartenenti allo stesso gruppo. All'aumentare della curvatura è necessario, per ottenere un grouping corretto, aumentare il coefficiente di diffusione, cioè la larghezza del nucleo: in questo modo segmenti disposti lungo una curva possono essere raggruppati nello stesso cluster. Per ottenere lo stesso risultato, occorre diminuire il raggio del nucleo, per analizzare al meglio lo stimolo.

Un altro esempio di interesse è rappresentato dal seguente stimolo:

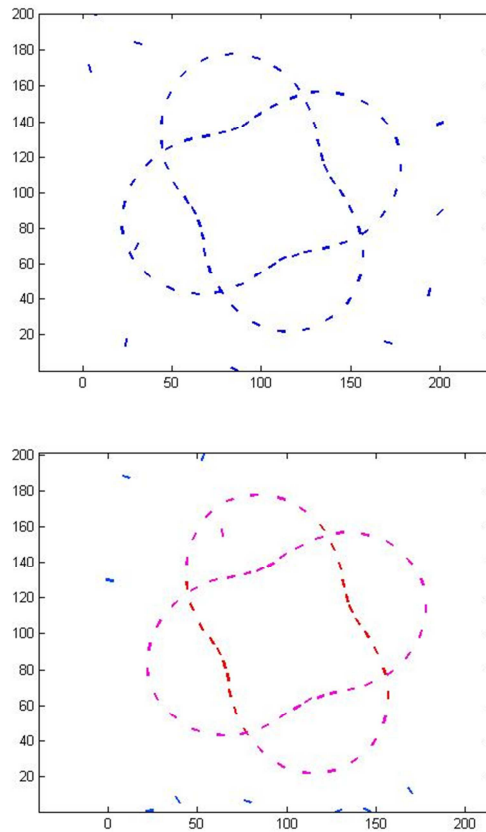


Figura 4.32:

Esempio di stimolo composto da due cerchi deformati (in alto) e risultato di clustering (in basso).

In questo stimolo, sono presenti due unità percettive che sono due cerchi deformati: i punti appartenenti a una singola unità non hanno curvatura costante ma variabile in base alla posizione del segmento sul contorno.

Come mostrano le immagini, non si ottengono tre gruppi perfettamente separati, come ci si aspetterebbe dall'esempio in esame.

In questo caso le informazioni sullo stimolo, coordinate x , y , Θ , sono insufficienti a ottenere una corretta segmentazione dello stesso. Una possibile soluzione per ottenere un buon clustering potrebbe essere l'introduzione di altre variabili, quali la velocità locale dei segmenti dello stimolo. È stato infatti dimostrato che cellule semplici in V1 sono distribuite in corteccia in modo tale da misurare ottimamente, oltre all'orientazione [4], anche proprio la velocità locale [2].

Conclusioni

L'obiettivo di questa tesi è stato quello di analizzare nel dettaglio modelli geometrici di percezione visiva, attraverso l'utilizzo di tecniche di riduzione di dimensionalità e di clustering. Per arrivare a questo si è iniziato, dopo una descrizione anatomica e fisiologica dell'occhio, con l'introduzione dei campi recettivi e dei filtri di Gabor, come filtri che descrivono i profili recettori di cellule semplici. È stata poi analizzata l'organizzazione della corteccia visiva e la sua struttura ipercolonnare, che organizza le cellule corticali in queste colonne in dipendenza dal parametro di orientazione.

Dopo aver introdotto i principi della Gestalt, è stata presentata la teoria dei campi di associazione visivi di Field, Hayes ed Hess, descrivendo quali caratteristiche due segmenti devono avere per poter essere associati allo stesso gruppo percettivo. Questi campi di associazione possono essere modellati mediante l'utilizzo del sistema di equazioni differenziali proposto da Citti-Sarti, le cui soluzioni sono curve integrali adatte a ricostruirne la struttura.

È stata descritta la connettività corticale tra neuroni, per comprendere come cellule appartenenti alla corteccia visiva primaria siano in grado di realizzare i campi di associazione, presentando poi il concetto di nucleo stocastico come modello di tale connettività e di densità di probabilità associata, per modellare l'attività di neuroni corticali. Sono stati successivamente descritti e analizzati questi nuclei al variare di determinati parametri, quali numero di percorsi, numero di passi utilizzati in ognuno degli stessi e costante di diffusione σ^2 , riportando esempi.

Infine, è stato presentato il fenomeno percettivo del grouping visivo, come risultato di un procedimento di riduzione di dimensionalità, attraverso gli algoritmi del K-means e dello Spectral clustering, dopo aver introdotto la teoria dei grafi. È stato illustrato il concetto di affinità tra stimoli e la costruzione della corrispondente matrice di affinità al variare dei parametri del nucleo, riportando diversi casi. Per comprendere più nel dettaglio la

possibilità di poter riprodurre, attraverso l'analisi spettrale di matrici di affinità calcolate utilizzando i modelli geometrici di Citti-Sarti, il fenomeno percettivo di grouping nello spazio $\mathbb{R}^2 \times S^1$ sono stati analizzati esempi applicativi, quali la segmentazione oggetto/sfondo e l'estrazione di oggetti con tecniche di riduzione di dimensionalità. Attraverso i primi esempi, rappresentati da un disco nero su sfondo bianco e una forma bianca con angoli su sfondo nero, si è mostrato uno dei fondamentali processi che stanno alla base di ogni esperienza visiva, l'articolazione figura-sfondo, che rappresenta la tendenza del campo recettivo a dividersi in ciò che è primario (l'oggetto in esame) da ciò che è secondario (lo sfondo). Attraverso l'estrazione di oggetti attraverso metodi di riduzione di dimensionalità, invece, sono stati analizzati i parametri che influenzano il fenomeno di grouping, mostrando quali permettono di partizionare bene i dati in cluster distinti.

Numerosi potrebbero essere gli sviluppi futuri che il lavoro svolto in questa tesi può fornire. A partire da quest'ultimo esempio presentato, sarebbe infatti interessante provare a effettuare la stessa analisi parametrica utilizzando stimoli più complessi, trovando così quali altri parametri dovrebbero intervenire per poter ottenere un buon clustering. Inoltre, sarebbe interessante valutare la segmentazione oggetto/sfondo con altri esempi, per poter così riuscire a distinguere cosa è foreground da cosa è background e comprendere al meglio questo fenomeno così studiato anche a livello psicologico.

Bibliografia

- [1] Adelson E. H., Bergen J. R., (1985), *Spatiotemporal energy models for the perception of motion*. Journal of the Optical Society of America 2: 284-299.
- [2] Barbieri D., Citti G., Cocci G., Sarti A. (2013). *A cortical-inspired geometry for contour perception and motion integration*. Preprint.
- [3] Bart M. ter Haar Romeny (2003). *Front- End Vision and Multi-Scale Image Analysis*. Computational Imaging and Vision 27: 150-90.
- [4] Bosking W. H., Zhang Y., Schofield B., Fitzpatrick D., (1997). *Orientation Selectivity and the Arrangement of Horizontal Connections in Tree Shrew Striate Cortex*. The Journal of Neuroscience 17(6): 2112–2127.
- [5] Citti G., Sarti A., (2006). *A cortical based model of perceptual completion in the roto-translation space*. Journal of Mathematical Imaging and Vision 24: 307-326.
- [6] Coifman R.R., Lafon S., (2006). *Diffusion maps*. Applied and Computational Harmonic Analysis.
- [7] Daugman J. G., (1985). *Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters*. Journal Optical Society of America 36: 1160-1169.
- [8] DeAngelis G. C., Ohzawa I., Freeman R. D., (1995). *Receptive-field dynamics in the central visual pathways*. Trends Neuroscience 18: 451-458.

[9] Field D., Hayes A., Hess R. F., (1993). *Contour integration by the human visual system: evidence for a local "Association Field"*. Vision Research 33: 173-179.

[10] Freeman W.T., Perona P., (1988). *A factorization approach to grouping*. Computer Vision 1406: 655-670.

[11] Gabor D. (1946). *Theory of communication*. IEE 93: 429-459.

[12] Gilbert C.D., Wiesel T.N. (1989). *Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex*. The Journal of Neuroscience 9: 2432–2442.

[13] Hubel D. H., Wiesel T. N., (1965). *Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat*. Journal of Physiology 28: 229-289.

[14] Hubel D. H., Wiesel T. N., (1976). *Functional architecture of macaque monkey visual cortex*. Royal Society of London B. 198: 1-59.

[15] Hubel D. H., (1995). *Eye, brain, and vision*.

[16] Jacobs D.W., Williams L.R. (1997), *Stochastic Completion Fields: A Neural Model of Illusory Contour Shape and Saliency*. Neural Computation 9(4):837-58.

[17] Jones J. P., Palmer L. A., (1987). *An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex*. Journal of Neurophysiology 58: 1233-1258.

[18] Kanisza G.,(1980). *Grammatica del vedere*. Il mulino, Bologna.

- [19] Kaplan E, Shapley R M (1986). *The primate retina contains two types of ganglion cells, with high and low contrast sensitivity*. PNAS 83 : 2755-2757.
- [20] Lafon, S. (2004). *Diffusion maps and geometric harmonics*. PhD Thesis, Yale University.
- [21] Lourens T., Wurtz R.P., (2000). *Corner detection in color images through a multiscale combination of end-stopped cortical cells*. Image and Vision Computing 18:531-541.
- [22] Malik J., Shi J., (2000). *Normalized cuts and image segmentation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22 (8), 888 – 905.
- [23] Malach R., Amir Y., Harel M., Grinvald A. (1993). *Relationship between intrinsic connections and functional architecture revealed by optical imaging and in vivo targeted biocytin injections in primate striate cortex*. PNAS 90:10469 –10473.
- [24] Meila, M. and Shi, J. (2001). *A random walks view of spectral segmentation*. In 8th International Workshop on Artificial Intelligence and Statistics (AISTATS).
- [25] Mumford D, (1994). *Elastica and computer vision*. Springer :491-506
- [26] Ringach D. L., Bredfeldt C. E., (2002). *Dynamics of spatial frequency tuning in macaque V1*. J Neuroscience 22(5): 1976-1984.

[27] Rock I., (1991). *L'eredità della psicologia della Gestalt*. Le Scienze 270.

[28] Von Luxburg U., (2006). *A tutorial on spectral clustering*. Technical Report No. TR-149.

[29] Weiss Y., (1999). *Segmentation using eigenvectors: a unifying view*. Computer Vision, (2):975-982.

[30] Wertheimer M., (1912). *Experimental Studies of the Perception of Motion*. Journal of Physiology 61: 161-265.

Ringraziamenti

Il mio primo ringraziamento non può che andare alla mia famiglia, in particolare ai miei genitori che mi hanno sempre aiutato nei momenti di difficoltà che questo percorso ha presentato, appoggiando le mie scelte e standomi vicino, facendomi affrontare il tutto serenamente, nonostante il mio caratteraccio.

Grazie infinite perché senza di voi tutto questo non sarebbe mai stato possibile (avrei probabilmente mollato molto prima, come ben sapete). Ringrazio anche i miei fratelloni, che mi hanno sempre incoraggiata e supportata.

I miei più sentiti ringraziamenti vanno a chi mi ha seguito in questo percorso, a partire dal Prof. Claudio Lamberti, che ha dimostrato in questi anni di avere fiducia in me, al Prof. Alessandro Sarti, per avermi dato l'opportunità di scoprire attraverso questa tesi un mondo nuovo e affascinante, e al Dott. Giacomo Cocci, per la sua disponibilità e chiarezza, per avermi seguito e aiutato con tanta pazienza durante lo svolgimento del lavoro, permettendomi di arrivare a tutto questo.

Ringrazio gli amici, delle diverse vallate: a partire da 'le amiche di Santa' Cate, Ceci, Ary, Marghe, Benny, che nel corso di questi anni mi sono state vicino, cercando di vedere sempre i lati positivi anche nei momenti più bui; 'le amiche di Bagno' Vale, Babi, Meli, Enri, Laura che hanno sopportato le numerose assenze (giustificate!) e mi hanno spronato a dare il massimo; 'gli amici dell'Università' Claudia, Silvia, Fanta, i milanesi, Micol (sarei sempre considerata un Ing.), Robi, Bene, che ho avuto il piacere di conoscere lungo questo percorso e hanno reso Ingegneria più piacevole e umana; 'gli amici di Cesena' Sassi, Chiara, Giulia, Marti, Valeria, Dani, Vale, Silvia con cui ho vissuto momenti stupendi e che ho avuto la fortuna di incontrare in questi ultimi anni.

L'ultimo ringraziamento, forse il più sentito va proprio a Ingegneria, perché nonostante a volte l'abbia odiata profondamente mi ha permesso di

incontrare il mio Luca. Grazie a Te e alla tua famiglia, per avermi accolto come una figlia, aiutato e soprattutto sopportato in questo ultimo periodo. Sono felice di raggiungere questo ulteriore traguardo, come sempre insieme.