

# Introduzione

Lo scopo di questa tesi è quello di presentare l'applicazione di tecniche legate alla Teoria di Taglia a un problema di analisi di immagini biomediche. Il lavoro nasce dalla collaborazione del gruppo di Matematica della Visione dell'Università di Bologna, con il progetto PERFECT del Centro di Ricerca ARCES, cui il gruppo stesso afferisce.

Il progetto PERFECT aspira a fornire un metodo semi-automatico di analisi di immagini perfusionali su tessuti tumorali. Per natura dei tessuti e difficoltà di acquisizione, i dati presentano un deciso rumore che compromette lo studio perfusionale secondo tecniche tradizionali.

La tesi si pone quindi come analisi preliminare di approccio alternativo ai metodi preesistenti.

I metodi sono principalmente di Topologia Algebrica Applicata, ambito emergente e rilevante nel mondo della Matematica Computazionale. Il nucleo dell'elaborazione è costituito da l'analisi di forma dei dati per mezzo di Funzioni di Taglia. Questa nozione è stata introdotta nel 1999 da Patrizio Frosini [1] e in seguito sviluppata principalmente dallo stesso, Massimo Ferri, Claudia Landi e altri collaboratori appartenuti al Gruppo di Ricerca di Matematica della Visione dell'Università di Bologna.

Nel Capitolo 1 esporrò brevemente alcune nozioni riguardanti l'ambiente in cui si sono svolte le analisi: alcune nozioni sulle neoplasie, i metodi diagnostici e d'acquisizione utilizzati.

Nel Capitolo 2 tratterò delle conoscenze teoriche necessarie per la costruzione dell'elaborazione.

Il Capitolo 3 riguarda la strutturazione dell'algoritmo di analisi e per finire nell'ultimo Capitolo mostrerò i risultati delle sperimentazioni svolte.



# Capitolo 1

## Premesse

Lo sviluppo di nuove terapie antitumorali costituisce uno dei maggiori ambiti di ricerca in campo medico. I miglioramenti volgono verso una maggiore efficacia e una azione meno invasiva delle cure nel tentativo di salvaguardare i tessuti sani.

Tra le analisi emergenti compare la Tomografia Computerizzata perfusionale, una tecnica di imaging radiologico che permette di esaminare le capacità funzionali del tessuto. Questo primo capitolo ha lo scopo di introdurre le premesse mediche e le tecniche di analisi finora utilizzate. Descriverò in particolare nel primo paragrafo alcuni aspetti dei tumori, in seguito la tomografia computerizzata e infine le tecniche di acquisizione e le tecniche di elaborazione dati attuate nel progetto Perfect.

### 1.1 Tumori

Il cancro è una delle maggiori cause di morte nel mondo: nel 2012 sono state contate più di 8 milioni di morti. Il tumore, o *neoplasia*, è definita come una massa di tessuto anomala con crescita irregolare e non controllata che persiste anche dopo che sia svanita la causa originante. E' una malattia che può affliggere diverse parti del corpo e manifestarsi in modi differenti.

La neoplasia si origina da uno sconvolgimento morfologico o funzionale di poche cellule. Queste cellule difettose, se l'organismo non è in grado di controllarle, iniziano un ripetuto processo di moltiplicazione che porta a un ingrandimento della massa atipica.

La neoplasia può acquisire una certa autonomia e modificare se stessa e i tes-

suti circostanti in funzione di un maggiore sviluppo. In questo processo può acquisire la capacità di attuare l'*angiogenesi*, cioè la creazione di nuovi vasi sanguigni, al fine di apportare un maggiore nutrimento alle cellule tumorali e favorirne la moltiplicazione. È una fase determinante, perchè in assenza di questa facoltà la dimensione si mantiene limitata a qualche millimetro di diametro.

Nella diagnosi delle neoplasie, quindi, la distinzione della presenza di nuovi vasi è un importante indicatore dello stato di crescita.

## 1.2 Diagnosi

La *Tomografia Computerizzata perfusionale* (TCp) è una tecnica diagnostica che combina la radiografia con lo studio perfusionale dei tessuti. La Tomografia Computerizzata è una pratica comunemente usata nello studio delle strutture anatomiche interne al corpo. La procedura prevede che il paziente venga disteso su un lettino e scansionato da un anello rotante in cui sono disposti in posizioni antpodali l'emettitore di raggi X e i detettori. L'esame può essere effettuato in due differenti modi, la scansione *assiale* (TAC) e la scansione *a spirale*. Nella prima il lettino si muove lungo l'asse dell'anello detettore e vengono acquisite delle sezioni, o *slice*, di corpo a intervalli di tempo regolari, una per ciascuna posizione del lettino. Nella scansione *a spirale* l'anello scannerizza il corpo compiendo un moto roto-traslatorio attorno al lettino. In questo caso viene acquisito tutto il volume in assenza di discontinuità.

La perfusione è il passaggio di fluido attraverso i vasi capillari in un tessuto. Quantitativamente è definita come il trasporto di sangue in un'unità di volume per unità di tempo. Una valutazione di questa quantità può dare un'indicazione della differente densità del tessuto.

I parametri principali sono:

- $BF$ , il flusso sanguigno. Un elevato valore indica la possibilità di un'irrorazione anomala;
- $BV$ , il volume di sangue che affluisce al tessuto in esame. Anche per questa grandezza, valori maggiori possono indicare una maggiore presenza di vasi e quindi suggerire un processo di angiogenesi;

- *MTT*, tempo di transito medio, indica il tempo medio che il sangue impiega ad attraversare una porzione delimitata da un'arteria o una vena. Contariamente ai precedenti un minore tempo indica una presenza maggiore di vasi;
- *PS*, superficie di permeabilità, calcola il prodotto tra permeabilità e la superficie totale interna dei vasi. Misura il flusso in uscita dal vaso verso il tessuto interstiziale. L'aumento di permeabilità riflette la maggiore permeabilità caratteristica dei nuovi vasi.

La procedura della TCp prevede l'iniezione nel paziente di un mezzo di contrasto; il liquido ha una differente capacità di assorbimento delle radiazioni e quindi nella scansione tomografica viene rilevato in modo differente. Valutando la concentrazione del mezzo di contrasto al variare del tempo e dello spazio si possono dedurre i parametri perfusionali e la struttura vascolare del tessuto. Ciò rende la TCp molto utile nell'individuare la presenza di neovascolarizzazione del tumore sia a scopo preventivo nell'evoluzione della neoplasia, sia nella valutazione dell'effetto di somministrazione di farmaci anti-angiogenetici.

L'utilizzo della TCp nello studio angiogenetico dei tumori è stato proposto per la prima volta da Axel nel 1980. Da allora le tecnologie hanno portato a notevoli miglioramenti in termini di qualità delle acquisizioni, ma anche di benessere del paziente con l'ottenimento della riduzione delle emissioni di radiazioni. Tuttavia sono presenti ancora numerosi problemi che rendono la tecnica oggetto di ricerca. Molti sono i problemi da cui è affetto: l'assenza di una procedura standard non permette di ottenere una riproducibilità affidabile nemmeno in esami con lo stesso operatore. Per questo è complicato anche valutare quali siano i parametri perfusionali più importanti.

Gli altri problemi legati a questa tecnica sono legati perlopiù a errori di acquisizione. Infatti sebbene sia chiesto al paziente di trattenere il respiro nel corso dell'esame non si possono prevenire altri movimenti esterni, né quelli degli organi interni, come quelli dovuti al battito cardiaco o a contrazioni muscolari involontarie. Inoltre la digitalizzazione dell'immagine radiologica effettua una rigida suddivisione che non tiene conto delle caratteristiche del tessuto e lo stesso pixel in slice successive potrebbe non definire interamente o parzialmente la stessa porzione.

A causa dell'effetto nocivo delle radiazioni non è inoltre possibile sottoporre il

paziente a troppe scansioni consecutive e ciò impedisce una regolarizzazione statistica dei dati.

### 1.3 Acquisizione dati e modelli

Il primo passo nell'acquisizione di dati da TCp è la scelta della zona da analizzare. Le prime scansioni vengono effettuate a lettino fermo e servono a individuare la regione lesionata. Un medico esperto traccia il contorno della regione interessata (ROI, '*Region Of Interest*') e di una zona interna a un vaso sanguigno. La prima è la fonte principale delle analisi e deve essere caratterizzata dall'assenza di grandi vasi, tessuti adiposi o zone d'aria. La seconda serve a determinare l'input arterioso necessaria per standardizzare l'analisi rispetto alla variabilità inter-paziente.

Il posizionamento della ROI costituisce un'altra fonte di problemi. Per quella tumorale non è stato ancora deciso su quale lato del tumore debba essere posizionata, mentre per quella arteriosa è incerto se sia preferibile un vaso di grandi o medie dimensioni. Essendo uno studio prolungato nel tempo, occorre considerare anche l'evoluzione della regione nel tempo. Questa valutazione è chiaramente molto affetta da errori dovuti ai movimenti. Una prima soluzione è quella di far determinare al medico la ROI per ogni slice (metodo *Manual*). La richiesta di un intervento umano allontana il processo dall'automatizzazione e fornisce contorni non facilmente sovrapponibili, ma è comunque preferibile a lasciare fissa la maschera nel tempo (metodo *Fixed*) in quanto comporterebbe l'inclusione nelle analisi di dati relativi a tessuti non attinenti allo studio. Il CVG (Computer Vision Group) di Bologna ha tentato di ovviare al problema sottoponendo i dati a un procedimento di allineamento automatico definito *best track*. Questo metodo prevede la scelta da parte del radiologo di una slice di riferimento, in cui è più evidente la porzione tissutale soggetta ad analisi, sulla quale viene tracciato il contorno della ROI. In seguito, viene determinato il centroide e traslato secondo gli spostamenti del tessuto, mantenendo però forma e dimensione costante.

Concluso questo passaggio, vengono estrapolati con software appositi i valori di concentrazione assunti ad ogni istante di tempo determinato. I valori di ciascun pixel possono essere visti come osservazioni e riportati come ordinate di in un grafico bidimensionale in cui l'ascissa è il tempo. La curva interpolante è detta Time Concentration Curve (TCC) ed è formata da tre stadi. La prima è detta *baseline* ed è il momento prima che il mezzo di contrasto

raggiunga l'area in questione. La seconda è l'*enhancement* e riguarda l'arrivo del tracciante caratterizzata da una forte crescita fino ad arrivare alla terza fase, l'*outflow* in cui fuoriesce dal tessuto. Queste tre fasi costituiscono il periodo di *first pass*. Successivo a questo c'è la fase interstiziale, quando il mezzo entrato in ricircolo, ripassa per il tessuto, verificando un nuovo incremento della concentrazione.

La fase di *first pass* è ben descritta dalla funzione Gamma-variato se si considera anche la fase di outflow. Si può però assumere che questa sia meno rilevante rispetto alla fase di introduzione del mezzo di contrasto. A fronte di questa supposizione si preferisce adottare il Modello di Hill, già ampiamente usato in farmacodinamica. L'equazione di Hill è la seguente [8]:

$$y(t) = E_0 + \frac{(E_{MAX} - E_0)}{\left(\frac{EC_{50}}{t}\right)^\alpha}$$

dove  $y(t)$  indica la concentrazione del liquido tracciante nel tempo,  $E_0$  il valore iniziale di baseline,  $E_{MAX}$  il valore di saturazione del tessuto,  $EC_{50}$  il tempo in cui la concentrazione è metà dell'intervallo  $[E_0, E_{MAX}]$ . La variabile  $\alpha$  è la pendenza della curva in fase di crescita.

Inizialmente il fitting veniva effettuato secondo un modello lineare basato solo su tre punti. Di certo costituiva un approccio più semplice perchè permetteva di ottenere velocemente il valore di pendenza, ma era più sensibile al rumore e incompatibile con l'uso di parametri di *goodness-of-fit* per lo studio dell'efficienza. Al contrario l'equazione di Hill, facendo uso di un insieme più grande di dati, ha un potere descrittivo maggiore e preserva meglio il trend delle TCCs. In questo modo acquisisce l'accuratezza e la robustezza.

Come però visto in precedenza i movimenti creano delle incoerenze tra pixel e può quindi capitare che la zona nella maschera relativa a un pixel non corrisponda alla stessa zona del tessuto. Questo rende il fitting difficile da effettuare, soprattutto nella scelta di quali osservazioni sono da ritenere buone in base all'informazione che vogliamo raccogliere. Per questo il CVG ha proposto un metodo, denominato *Look and Remove*(LAR), basato un algoritmo RANSAC. La procedura di fitting viene effettuata su un campionamento di dati scelto in modo casuale e valutato con un indice di confidenza. In modo iterativo vengono scelti i campioni con indice maggiore fino ad ottenere una accuratezza del 90%. Questo metodo è tuttora in fase di sperimentazione all'interno di un progetto di dottorato che prevede la disposizione di linee guida per l'intero protocollo.





## Capitolo 2

# Impianto teorico-matematico

L'aspetto più innovativo di questo lavoro è costituito dal tipo di approccio attuato. Il nucleo costituente è la Teoria della Taglia introdotta nel 1990 da P. Frosini [1] per distinguere varietà differenziabili a meno di classi di similitudine. L'obiettivo, allora, era trovare una distanza nello spazio delle varietà  $C^\infty$ -differenziabili che si annullasse quando due varietà differivano per una trasformazione di similitudine. La difficoltà del calcolo di tale distanza ha portato a sviluppare un descrittore di forma e una seconda distanza che approssimasse la prima.

Questo procedimento di semplificazione ha reso accessibile la teoria alle applicazioni [2],[4][6][7], primo fra tutti il riconoscimento di forme e analisi immagini ???. Ciò che più caratterizza questi strumenti è la proprietà di invarianza rispetto a trasformazioni e una certa resistenza al rumore e rendendola una valida alternativa ai metodi precedentemente usati.

In questo capitolo introduco la struttura matematica su cui si fondano i metodi usati nell'analisi.

Nel primo paragrafo tratterò della Teoria della Funzione di Taglia nelle sue definizioni e teoremi generali, mentre nel secondo tratterò l'aspetto computazionale.

Successivamente tratterò brevemente di altri strumenti usati soprattutto in problemi di Feature Extraction e Machine Learning. Per primo, la riduzione multidimensionale: essa prevede la semplificazione di uno spazio di parametri di dimensione elevata in uno a dimensione più bassa, favorendo la visualizzazione e accentuando le caratteristiche principali.

Infine la clusterizzazione, che è una tecnica che si occupa di dividere un insieme di dati in gruppi secondo criteri di similarità e fornisce una possibilità

di classificazione.

## 2.1 Teoria della funzione di taglia

Chiamiamo *coppia di taglia* la coppia ordinata  $(M, \phi)$ , dove  $M$  è uno spazio topologico e  $\phi : X \rightarrow \mathbf{R}$  è una funzione continua, detta *funzione misurante*. Assumiamo che  $M$  sia uno spazio topologico di Hausdorff, non vuoto, compatto, connesso e localmente connesso.

Costruiamo ora una distanza sullo spazio delle coppie, tale da distinguere se due forme siano simili.

**Definizione 2.1.** *Date due coppie di taglia  $(M, \phi)$ ,  $(N, \psi)$  e  $H(M, N)$  lo spazio degli omeomorfismi da  $M$  a  $N$ , chiamiamo **misura naturale di taglia** la funzione:*

$$\begin{aligned} \Theta : H(M, N) &\longrightarrow \mathbf{R} \\ f &\longmapsto \max_{P \in M} |\phi(P) - \psi(f(P))| \end{aligned}$$

Questa funzione conta il massimo scarto tra le funzioni misuranti nel passaggio da  $M$  a  $N$  tramite la deformazione omeomorfa  $f$ .

Considerando l'estremo inferiore sugli omeomorfismi da  $M$  a  $N$  della funzione  $\Theta$  otteniamo una pseudodistanza sullo spazio delle coppie di taglia:

**Definizione 2.2.** *Fissati  $(M, \phi)$ ,  $(N, \psi)$  coppie di taglia*

$$\Sigma[(M, \phi), (N, \psi)] = \begin{cases} +\infty & \text{se } H(M, N) = \emptyset \\ \inf_{f \in H(M, N)} \{\max_{P \in M} |\phi(P) - \psi(f(P))|\} & \text{altrimenti} \end{cases}$$

Se introduciamo sullo spazio delle coppie di taglia una relazione d'equivalenza  $\simeq$  così definita:

$$M \simeq N \iff \Sigma[(M, \phi), (N, \psi)] = 0$$

induciamo una metrica  $\sigma$  sullo spazio quoziente, chiamata *distanza naturale di taglia*.

Vorrei far notare che la differenza fra i due spazi, o forme, è mediata dai valori delle funzioni misuranti. Diverse funzioni misuranti comportano un diverso valore di  $\Theta$ , quindi è di grande interesse associare la giusta funzione misurante agli spazi. Per definizione la distanza naturale di taglia è difficile da calcolare perchè prende in considerazione tutti gli omomorfismi tra i due spazi.

La teoria allora prevede altri due strumenti: le *funzioni di taglia* e la *distanza di matching*. La funzione di taglia di una coppia  $(M, \varphi)$  intuitivamente tiene conto della variazione della struttura topologica degli insiemi di sottolivello di  $\varphi$ . Prima di definirla formalmente occorre definire dei concetti preliminari.

**Definizione 2.3.** *Data una coppia di taglia  $(M, \varphi)$  e  $x \in \mathbf{R}$ , sia  $M\langle\varphi \leq x\rangle$  l'insieme di sottolivello  $\{P \in M : \varphi(P) \leq x\}$ . Per ogni  $y \in \mathbf{R}$  e  $P, Q \in M$ , diremo che  $P$  e  $Q$  sono  $(\varphi \leq y)$ -omotopi se esiste un cammino  $\gamma : [0, 1] \rightarrow M$  tale che  $\gamma(0) = P$ ,  $\gamma(1) = Q$  e  $\varphi(\gamma(\tau)) \leq y$  per ogni  $\tau \in [0, 1]$*

Questa definizione definisce una relazione d'equivalenza sull'insieme di sottolivello  $M\langle\varphi \leq x\rangle$ . Poniamo  $\Delta^+ = \{(x, y) \in \mathbf{R}^2, x < y\}$  e  $\Delta = \{(x, y) \in \mathbf{R}^2, x = y\}$ .

**Definizione 2.4.** *La funzione di taglia di  $(M, \varphi)$  è una funzione  $l_{(M, \varphi)} : \Delta^+ \rightarrow \mathbf{Z}$  che associa ad ogni coppia  $(x, y)$  il numero delle classi di equivalenza in cui  $M\langle\varphi \leq x\rangle$  è diviso rispetto alla relazione di  $(\varphi \leq y)$ -omotopia.*

Per queste funzioni valgono le seguenti proprietà:

*Osservazione 2.1.1.*  $l_{(M, \varphi)}^*(x, y)$  è non decrescente in  $x$  e non crescente in  $y$ ;

*Osservazione 2.1.2.*  $l_{(M, \varphi)}^*(x, y)$  è finita per  $x < y$ .

*Osservazione 2.1.3.*  $l_{(M, \varphi)}^*(x, y) = 0$  per ogni  $x < \min_{P \in M} \varphi(P)$

*Osservazione 2.1.4.*  $l_{(M, \varphi)}^*(x, y)$  è costante e uguale al numero delle componenti connesse per archi di  $M$  per ogni  $x, y > \max_{P \in M} \varphi(P)$

*Osservazione 2.1.5.*  $l_{(M, \varphi)}^*(x, y) = \infty$  per ogni  $x, y \in \mathbf{R}$  per i quali esiste un  $Q \in M$ , non isolato, per cui  $x > \varphi(Q)$  e  $y < \varphi(Q)$

Definiamo ora delle funzioni analoghe, ma richiedono la connessione classica invece della connessione per archi.

**Definizione 2.5.** *Data una coppia di taglia  $(M, \varphi)$  e  $x \in \mathbf{R}$ , sia  $M\langle\varphi \leq x\rangle$  l'insieme di sottolivello  $\{P \in M : \varphi(P) \leq x\}$ . Per ogni  $y \in \mathbf{R}$  e  $P, Q \in M$ , diremo che  $P$  e  $Q$  sono  $\langle\varphi \leq y\rangle$ -connessi se esiste una componente connessa di  $M\langle\varphi \leq y\rangle$  che li contiene entrambi.*

Otteniamo un'altra relazione d'equivalenza sull'insieme di sottolivello  $M\langle\varphi \leq x\rangle$ .

**Definizione 2.6.** La funzione di taglia ridotta di  $(M, \varphi)$  è una funzione  $l_{(M, \varphi)}^* : \Delta^+ \rightarrow \mathbf{Z}$  che associa ad ogni coppia  $(x, y)$  il numero delle classi di equivalenza in cui  $M\langle \varphi \leq x \rangle$  è diviso rispetto alla relazione di  $\langle \varphi \leq y \rangle$ -connettività.

D'ora in poi considereremo solo le funzioni di taglia ridotte, perchè permettono di ottenere la continuità a destra rispetto alla variabile  $x$  che alla variabile  $y$ , non in generale vera per le funzioni di taglia classiche. Questa richiesta si fonda sul fatto che l'informazione di una funzione di taglia giace tutta in pochi punti, dei particolari punti di discontinuità della funzione.

**Definizione 2.7.** Per ogni punto  $p = (x, y) \in \Delta^+$  definiamo la molteplicità di  $p$  per  $l_{(M, \varphi)}^*$ ,  $\mu(p)$  come il minimo al variare di  $\epsilon$  numero reale positivo tale che  $x + \epsilon < y - \epsilon$  della quantità:

$$l_{(M, \varphi)}^*(x + \epsilon, y - \epsilon) - l_{(M, \varphi)}^*(x - \epsilon, y - \epsilon) - l_{(M, \varphi)}^*(x + \epsilon, y + \epsilon) + l_{(M, \varphi)}^*(x - \epsilon, y + \epsilon)$$

**Definizione 2.8.** Con le ipotesi precedenti, diremo che  $p$  è un cornerpoint proprio per  $l_{(M, \varphi)}^*$  se  $\mu(p) > 0$ .

Analogamente:

**Definizione 2.9.** Per ogni  $r$  retta verticale di equazione  $x = k$  definiamo la molteplicità di  $r$  per  $l_{(M, \varphi)}^*$ ,  $\mu(r)$  come il minimo al variare di  $\epsilon$  numero reale positivo tale che  $k + \epsilon < 1/\epsilon$  della quantità:

$$l_{(M, \varphi)}^*(k + \epsilon, 1/\epsilon) - l_{(M, \varphi)}^*(k - \epsilon, 1/\epsilon)$$

**Definizione 2.10.** Con le ipotesi precedenti, diremo che  $r$  è una cornerpoint all'infinito per  $l_{(M, \varphi)}^*$  se  $\mu(p) > 0$ . La identificheremo inoltre con la coppia  $(k, \infty)$

Le molteplicità sono sempre numeri positivi per l'Oss. 1.1. Possiamo notare che per l'ipotesi di connessione di  $M$  esiste sempre e un solo cornerpoint all'infinito di molteplicità 1. Si può però estendere la teoria a  $M$  formato da un numero finito di componenti connesse: in questo caso ci possono essere più cornerpoint all'infinito con molteplicità anche maggiore di 1.

Il seguente teorema conferma l'affermazione precedente che le funzioni di taglia dipendono dai cornerpoints propri.

**TEOREMA 2.1.1.** *Se  $\Delta^* := \Delta^+ \cup \{(k, \infty), k \in \mathbf{R}\}$ , per ogni  $(\bar{x}, \bar{y}) \in \Delta^*$  vale*

$$l_{(M,\varphi)}^*(\bar{x}, \bar{y}) = \sum_{\substack{(x,y) \in \Delta^* \\ x \leq \bar{x}, y > \bar{y}}} \mu((x, y))$$

Per una dimostrazione accurata del Teorema si rimanda a [3].

Il confronto tra due funzioni di taglia  $l_1$  e  $l_2$  può essere fatto sui rispettivi insiemi di cornerpoints. Sia  $C_i$  l'insieme dei cornerpoint di  $l_i$  contati con molteplicità unito alla retta diagonale  $\{(x, y) \in \mathbf{R}^2, x = y\}$ .

Dati  $(x, y)$  e  $(x', y')$  definiamo  $\delta((x, y), (x', y')) = \min \left\{ \max\{|x - x'|, |y - y'|\}, \max \left\{ \frac{y-x}{2}, \frac{y'-x'}{2} \right\} \right\}$

**Definizione 2.11.** *Nelle notazioni precedenti e posto  $\Pi$  l'insieme delle biezioni tra  $C_1$  e  $C_2$ , definiamo distanza di matching tra  $l_1$  e  $l_2$  il valore*

$$d_{match} = \min_{\pi \in \Pi} \max_{p \in C_1} \delta(p, \pi(p))$$

La distanza di matching calcola la minima spesa necessaria per portare i cornerpoint della prima funzione sui cornerpoint della seconda al variare le associazioni possibili tra i due insiemi. L'aggiunta dei punti della diagonale copre le differenze di cardinalità di cornerpoints e non porta costi aggiuntivi dato che lo spostamento tra due punti della diagonale è nullo per definizione.

Il primo risultato importante è che questa distanza da una stima della pseudo-distanza naturale.

**TEOREMA 2.1.2.** *Sia  $\epsilon \geq 0$  un numero reale,  $(M, \varphi)$ ,  $(N, \psi)$  due coppie di taglia con  $M$  e  $N$  omeomorfi e siano  $l_{(M,\varphi)}^*$  e  $l_{(N,\psi)}^*$  le funzioni di taglia ad esse associate. Allora*

$$d_{match}(l_{(M,\varphi)}^*, l_{(N,\psi)}^*) \leq \sigma[(M, \varphi), (N, \psi)]$$

L'altra proprietà chiave, che la rende un buono strumento per le applicazioni è la capacità di mantenersi stabile rispetto a piccole variazioni delle funzioni misuranti. In questo modo forme che differiscono per piccoli perturbazioni, rispetto alla lettura data dalla funzione misurante, hanno distanza piccola.

**TEOREMA 2.1.3.** (Teorema di stabilità) *Siano  $(M, \varphi)$  e  $(N, \phi)$  due coppie di taglia.*

$$\max_{P \in M} |\varphi(P) - \phi(P)| \leq \epsilon \implies d_{match}(l_{(M,\varphi)}^*, l_{(M,\phi)}^*) \leq \epsilon$$

La dimostrazione dettagliata è in [5].

## 2.2 Riduzione dimensionale

Con *scaling multidimensionale* (multidimensional scaling, MDS) si intendono una serie di tecniche di esplorazione dati che permette di visualizzare corrispondenze tra punti. Generalmente il Multidimensional Scaling parte da una matrice fatta di valori di caratteri ordinali, cioè di variabili il cui intervallo di valori possibili gode della proprietà di ordine totale; in seguito le trasforma in matrice con valori su scala di rapporti.

Il MDS non dispone gli elementi in modo esatto: a partire da una configurazione vengono effettuati dei passi successivi per ottenerne una nuova configurazione che corrisponda al meglio alla disposizione dei punti dettata dalla matrice di input [9]

Il metodo usato in questa tesi è il metodo di Sammon ed è un metodo non lineare, fa parte cioè di una categoria di metodi che hanno la caratteristica e lo scopo di mantenere la struttura non lineare dei dati di partenza. Compare per la prima volta nel 1989 in un articolo di Sammon come metodo di riduzione bidimensionale che tenta di ottenere una configurazione per cui sono preservate le distanze. Dato  $x_1, \dots, x_N$  un insieme punti appartenenti in uno spazio  $d$ -dimensionale e  $d(i, j)$  la matrice delle distanze reciproche, dove  $i, j = 1, \dots, N$ . Supponendo che esista una configurazione in uno spazio  $m$ -dimensionale con  $m < d$ , in cui ciascun punto  $x_i$  è associato a un punto della configurazione. Data  $D(i, j)$  la matrice delle distanze tra i punti corrispondenti a  $x_i$  e  $x_j$  in questa configurazione, Sammon definisce un errore tra i due insiemi:

$$E = \frac{1}{\sum_i \sum_{j, i < j} d(i, j)} \sum_i \sum_{j, i < j} \frac{[d(i, j) - D(i, j)]^2}{d(i, j)}$$

Questo valore è chiamato solitamente stress e l'algoritmo prevede di trovare, fissato  $m$ , una configurazione per cui  $E$  sia minimale. Il processo di minimizzazione venne originariamente effettuato tramite il metodo della discesa del gradiente, ma negli anni seguenti sono state proposte anche metodi alternativi (Kirkpatrick, 1984). L'algoritmo termina quando viene raggiunta la tolleranza richiesta o sull'errore o sulla differenza tra errori di iterazioni successive. Si può partire da una configurazione casuale iniziale oppure, per accelerare i tempi di esecuzione, partendo da una configurazione ottenuta con PCA (Principal component analysis).

## 2.3 Clustering

Il *clustering* è un'ulteriore tecnica di analisi esplorativa dei dati. Si distingue dal multidimensional scaling innanzitutto per lo scopo: mentre la prima mira a ottenere una visualizzazione da coefficienti di similarità, il clustering mira a individuare e raggruppare punti in un insieme di dati. Può essere utilizzato anche su insiemi di cui si vogliono verificare caratteristiche già conosciute, ma perlopiù è molto utile per individuare strutture intrinseche di dati vergini. Si suddivide in clustering gerarchico e partizionale. Il clustering gerarchico crea una successione di gruppi annidati e a sua volta è suddiviso in agglomerativo (Bottom-Up) e divisivo (Top-Down): i primi partono da gruppi piccoli che crescono e si unificano fino a diventare un unico corpo; i secondi attuano il procedimento inverso, da un corpo unico si separano dando via a insiemi sempre più piccoli. Qui è possibile vedere quali sono i cluster che ci mettono più tempo a accorparsi (o separarsi) e quindi decidere quali sono i principali.

In questa tesi è stato usato invece un metodo partizionale: il metodo *k-means* [10]. Per questo tipo di metodo viene generata una singola partizione cercando di evidenziare i gruppi preesistenti. L'algoritmo effettua una divisione dell'insieme  $\{x_i, i = 1, \dots, N\}$  in  $k$  classi  $P_1, \dots, P_k$ , minimizzando la funzione

$$E(A, C) = \sum_{j=1}^k \sum_{x_i \in P_j} \|x_i - c_j\|^2$$

dove  $c_j$  per  $j = 1, \dots, k$  sono i centroidi dei  $P_j$  e  $A$  è la matrice di verità  $k \times N$  che indica l'appartenenza del punto  $x_i$  al cluster  $j$ -esimo. Chiediamo che  $0 < k < N$  e che  $\{P_j, j = 1, \dots, k\}$  costituisca una partizione dell'insieme  $\{x_i, i = 1, \dots, N\}$  con  $P_j \neq \emptyset \quad \forall j = 1, \dots, k$ . Anche in questo caso il processo è iterativo e si può partire da una partizione casuale o da risultati comprovati.



# Capitolo 3

## Svolgimento dell'elaborazione

In questo capitolo mi accingo a descrivere l'implementazione dell'elaborazione dei dati svolta nella regione di interesse. Il mio intento principale è quello di confrontare ciascun pixel nel suo comportamento perfusionale per poter distinguere le diverse reazioni e individuare comportamenti simili, qualora ce ne siano.

Nei primi due paragrafi spiegherò come avviene il trattamento iniziale cui sono stati sottoposti i dati per poter applicare su di essi la teoria delle Funzioni di Taglia, trattata nel capitolo precedente. Per prima cosa occorre dunque interpretare il dato in senso topologico e in particolare come coppia di taglia. Parlerò quindi delle quattro interpretazioni di forma che ho scelto di utilizzare in questa analisi. Infine, negli ultimi due paragrafi, mi dedicherò alle diverse modalità di visualizzazione dei risultati: a una di queste in particolare dedicherò una specifica analisi per giudicarne la validità.

### 3.1 Scelta dello spazio topologico

Mi sono stati forniti, per ciascun esame, una matrice  $n \times m$ , dove  $n$  è il numero di pixel e  $m$  il numero di osservazioni. La riga  $k$ -esima corrisponde al vettore di perfusione del  $k$ -esimo pixel della ROI.

Analizzando il problema da un punto di vista fisiologico, partiamo dal fatto che non ci sono degli standard da seguire o dei campioni validi sempre. La manifestazione del tumore, infatti, è diversa da paziente a paziente, poiché diversa può essere la sua posizione ed estensione all'interno dell'organo, diversi i tipi di tessuti coinvolti. Inoltre, la massa tumorale stessa presenta

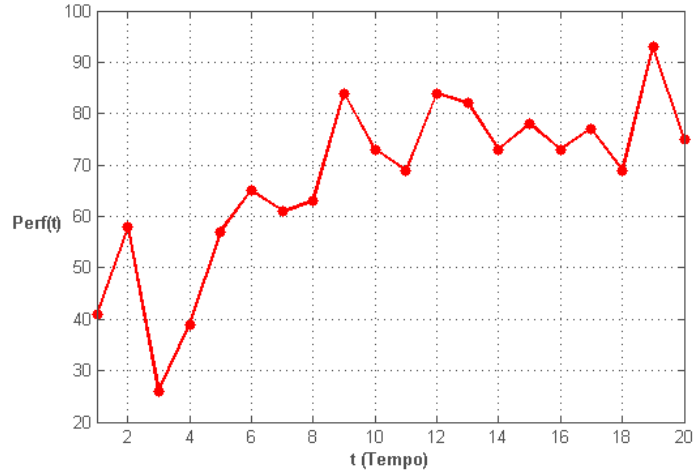


Figura 3.1: Intensità di perfusione di un singolo pixel

una variabilità dai caratteri non regolari. Partendo da queste constatazioni possiamo affermare che non è possibile dunque paragonare l'intera regione di interesse con un'altra analoga, poiché non esiste. Non possiamo dunque scegliere di associare un solo spazio topologico ad una intera regione.

La scelta da me intrapresa è quella di considerare un oggetto topologico per ogni singolo pixel. Inizialmente ho pensato di ampliare l'oggetto del quale prendo in considerazione la forma, in quanto la semplice retta spezzata con 20 nodi appariva un oggetto troppo scarno.

Per queste ragioni ho scelto di considerare come prima cosa, fissato un pixel, un reticolo  $N \times N$  corrispondente al rettangolo contenente il grafico del vettore di perfusione:

dato  $P$  il vettore  $1 \times m$  di perfusione, l'insieme dei vertici è costituito dall'insieme

$$\left\{ \left( s \frac{m}{N}, t \frac{A}{N} \right) \quad s, t = 1, \dots, N \right\}$$

con  $A = \max(P) - \min(P)$ ; i lati sono quelli relativi a un'adiacenza a quattro.

Questa costruzione riguarda le prime tre scelte, mentre la quarta è il risultato di una riconsiderazione della struttura più semplice, la retta spezzata. Questa analisi è stata fatta per completare l'analisi precedente e rivalutare l'assunzione precedente.

## 3.2 Scelte di funzione misurante

Volendomi occupare della lettura di forma secondo la Teoria della Taglia occorre aggiungere una funzione misurante allo spazio topologico scelto. La scelta di grafo porta a una definizione semplificata della funzione: basta definire i valori sui vertici che vengono poi estesi ai lati.

Per il caso del reticolo, ho scelto tre funzioni:

- funzione *distanza*: a ciascun vertice  $v = (v_x, v_y)$  associa la distanza dall'insieme dei dati,  $\min_{i=1, \dots, m} \|(v_x, v_y) - (i, Perf(i))\|_2$ ;
- funzione *variazione*: a ciascun vertice associa la distanza dall'insieme  $\{(i, \frac{Perf(i)-Perf(i-1)}{i}), \quad i = 2, \dots, 20\} \cup \{(1, Perf(1))\}$
- funzione *peso*: a ciascun vertice  $v = (v_x, v_y)$  associa la distanza dall'insieme dei dati  $\min_{i=1, \dots, m} \|(v_x, v_y) - (i, \frac{Perf(i)}{i})\|_2$

Ho deciso di prendere in considerazione la funzione *variazione* e la funzione *peso*, per valutare il comportamento del vettore in base al tempo, cosa che invece non emerge utilizzando la prima funzione. Nel caso della retta spezzata ho preferito scegliere la funzione *altezza*, la quale associa ad ogni vertice il valore di perfusione.

## 3.3 Confronto

Per ciascun pixel è calcolata la funzione di taglia e effettuato il confronto a due per ogni coppia di pixel. L'algoritmo utilizzato è quello esposto in ???. Il risultato di questo procedimento è una matrice  $n \times n$  delle distanze reciproche tra i pixel della ROI.

Questa operazione è la più costosa di tutto il procedimento, perché richiede due cicli for annidati per un totale di  $\frac{n(n-1)}{2}$  confronti.

## 3.4 Visualizzazione

A partire dalla matrice delle distanze si possono considerare diverse modalità di visualizzazione. La prima di queste è la modalità di visualizzazione che considera i pixel appartenenti ad uno spazio multidimensionale del

quale non si conoscono le coordinate, ma si conosce la rispettiva distanza tra i diversi pixel. Dal momento che ciascun pixel è definito da  $m$  parametri di osservazione, possiamo affermare che lo spazio in cui è inserito è  $m$ -dimensionale. Prendendo in considerazione gli esami effettuati, notiamo che il numero delle osservazioni  $m$  era 19 o 20. Tuttavia, questo dato ci porta ad escludere una visualizzazione diretta della nube di punti. Inoltre, un numero ridotto di osservazioni avrebbe costituito un set troppo ridotto per poterne discernere delle informazioni utili. Come prima cosa occorre dunque ridurre la dimensione dello spazio di giacenza dei pixel, servendosi di metodi di Multidimensional Scaling (MDS). Il requisito fondamentale è che tali metodi conservino quanto più possibile l'informazione di partenza, in modo tale che non vengano modificate le distanze tra un pixel e l'altro. Per questo motivo vengono preferiti metodi non lineari, come il metodo di Sammon, utilizzato per questo tipo di analisi. Una prima visualizzazione è costituita da un grafico a dispersione 3D, cioè l'insieme dei punti viene immerso in uno spazio tridimensionale. In più ho aggiunto un marcatore di colore che indica l'appartenenza del pixel ad un cluster. La clusterizzazione viene effettuata direttamente sulla matrice con un metodo *k-means*. In un secondo momento viene effettuata la proiezione su uno spazio bidimensionale al quale viene aggiunta come terza coordinata l'indice di clustering. Questo metodo permette una visualizzazione separata dei cluster su piani diversi, pur mantenendo anche una visualizzazione generale della nube di punti. Infine, la visualizzazione che, a mio parere, si rivela di più facile comprensione e valutazione, è quella in cui viene proposta una mappa colorimetrica della ROI. Questo tipo di rappresentazione ci permette, infatti, di evidenziare le zone del tessuto che hanno un comportamento diverso rispetto alla lettura svolta, analogamente alle colormap di perfusione, già ampiamente usate nelle analisi classiche. Questo modo è ottenuto associando a ciascun pixel un singolo valore: dalla matrice delle distanze di effettua una riduzione ad una dimensione; il vettore risultante vien in seguito disposto sulla maschera nella posizione corrispondente.

### 3.5 Analisi di qualità

A causa di una mancanza di un ground truth, cioè di un riferimento di validità dello studio, l'unica verifica che si può fare consiste nel determinare quale tipo di informazione contiene l'immagine ottenuta. Infatti, non è pos-

sibile determinare dal punto di vista qualitativo e quantitativo le regioni più o meno perfuse tramite un'analisi istologica dei tessuti o tramite altri mezzi. La stima di qualità alla quale sono pervenuto dopo queste attente analisi, è il risultato della valutazione di due indicatori: il *coefficiente di variazione medio* e l'*entropia locale media*. Entrambi i valori sono bene definiti a livello locale, ma ho voluto ottenere un indicatore che rappresentasse la maschera nel suo intero, facendo una media dei valori locali.

Data un'immagine essa è rappresentabile da una matrice  $L \times L$ . Per ogni pixel di posizione  $(i, j)$  considero la maschera locale  $M(i, j)$  di larghezza  $r$ . Il *coefficiente di variazione*  $CV(i, j)$  è il rapporto  $\frac{\sigma^{i,j}}{\mu^{i,j}}$  dove  $\sigma^{i,j}$  e  $\mu^{i,j}$  sono rispettivamente varianza e media della maschera  $M(i, j)$ . L'*entropia locale* invece è l'entropia della maschera calcolata con la funzione di Matlab `entropyfilt`.



# Capitolo 4

## Esiti delle sperimentazioni

Nel capitolo seguente mostro i risultati ottenuti dagli esperimenti effettuati. Ogni paragrafo descrive un esame differente: i primi due sono relativi a tumori localizzati nella regione polmonare, il terzo nel fegato. Per ciascuno dei tre casi ho analizzato sia i dati allienati manualmente, allineamento *Manual*, sia quelli trattati con allineamento *Best*.

Per la fase di riduzione dimensionale ho usato l'algoritmo `mdscale` con criterio di *Sammon*, condizione d'arresto posto a 1000 iterazioni massime o tolleranza di  $2e-3$ . Come dato iniziale ho usato quello di default per ogni funzione, eccetto che nella riduzione a una dimensione per qualche funzione con allineamento *Manual* per le quali avveniva un'unificazione di punti. Ciò costituiva la riduzione della cardinalità dell'insieme e non dava possibilità di ricondurre i punti ottenuti a quelli originari.

Le alternative provate sono: mantenere il criterio di *Sammon* e inserendo come dato iniziale una configurazione casuale; oppure adottare il criterio *strain*, un criterio classico nella riduzione dimensionale.

Per le colormap ottenute ho proposto un'analisi di sogliatura basata sull'istogramma, utilizzando la funzione di Matlab `multithresh` che calcola  $k$  livelli di soglia per mezzo del metodo di Otsu. Nella mia analisi ho posto  $k = 2$  e visualizzato le zone sottostanti il livello inferiore e sovrastanti quello superiore, fornendo inoltre per entrambi un calcolo del numero di pixel delle componenti connesse di cui erano composti.

Infine ho testato il numero di suddivisioni con cui effettuare la clusterizzazione, scegliendo 4 come valore ottimale.

## 4.1 Esame 1

Il primo esame riguarda immagini TC perfusionali di un paziente affetto da tumore nel polmone. È costituito da 20 slice temporali e la ROI è formata da 3385 pixels.

Per ogni funzione sono mostrate le immagini relative alla visualizzazione ottenuta con riduzione monodimensionale: da sinistra a destra le colonne riguardano l'allineamento *Best*, l'allineamento *Manual* con riduzione secondo il criterio di Sammon e ove calcolato l'allineamento *Manual* con riduzione secondo il criterio strain. Possiamo distinguere in tutte le immagini relative all'allineamento *Manual*, una zona particolarmente distinta dal resto della maschera. Essendo localizzata sul bordo superiore e laterale, ho giudicato la presenza di tale macchia come dovuta a un errato posizionamento della maschera. Questo errore introduce probabilmente dati relativi a una porzione di tessuto esterna alla lesione tumorale.

Si può notare inoltre che la presenza di una zona così atipica comprometta anche la relazione esistente tra i restanti pixels.

In questo caso suppongo che l'allineamento *best track* sia stato effettuato in modo adeguato, essendo le immagini della colonna di sinistra quasi privi di artefatti sul bordo.

Tra le quattro funzioni misuranti utilizzate sembra che la funzione *peso*, sia in grado di dare una lettura più significativa alla ROI, in quanto distingue aree con distribuzione più regolare rispetto alle altre.

Ricordo che il valore di ogni singolo pixel della colormap non riguarda una determinata grandezza ma è indice di una distanza tra pixel. I gradienti di colore sono stati scalati a due a due (associando le figure con stessa funzione misurante), ma generalmente differiscono al variare della funzione misurante. In 4.1 sono rappresentati gli indicatori di rumore definiti nel capitolo precedente. Questo grafico concorda con la supposizione che la funzione *peso* dia il risultato migliore in qualità. Per le immagini relative all'esame con allineamento *Manual* vengono calcolati valori relativamente più bassi a quelli con allineamento. Questo penso sia dovuto al fatto che, ad eccezione della funzione *altezza* con riduzione di Sammon caratterizzata dalla presenza di granulosità, le altre sono contraddistinte da una zona interna principalmente omogenea che con regolarità si collega alle zone atipiche posizionate sul bordo.



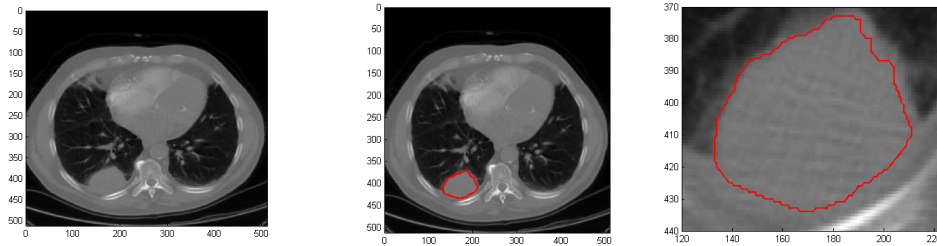


Figura 4.1: Da sinistra a destra un fotogramma completo, la maschera all'interno di esso e un ingrandimento della ROI

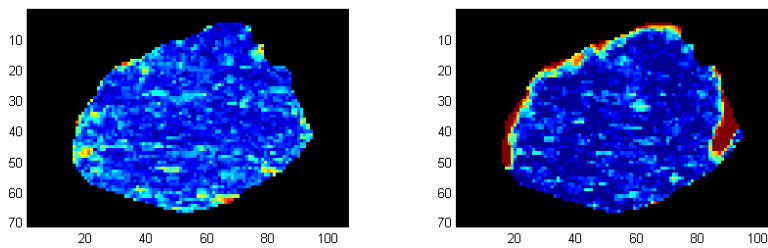


Figura 4.2: Esito della riduzione monodimensionale della funzione *distanza*, a sinistra con allineamento *Best*, a destra *Manual* e riduzione con criterio di Sammon

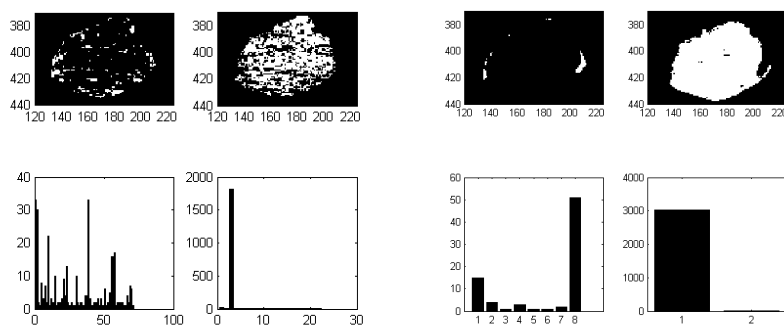
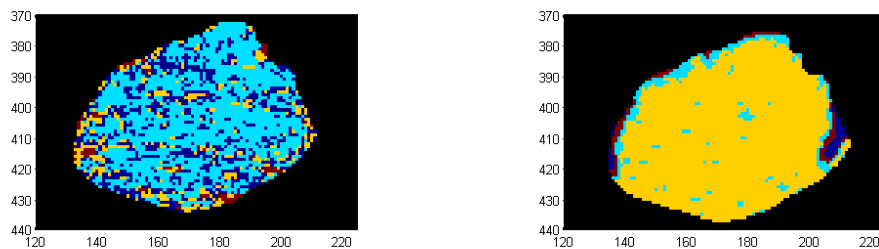
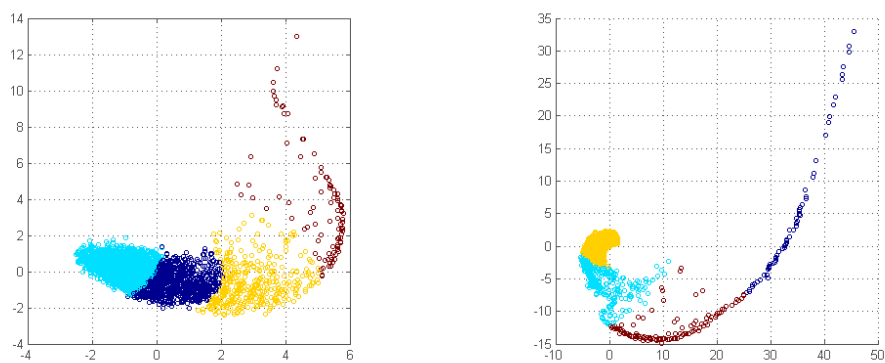
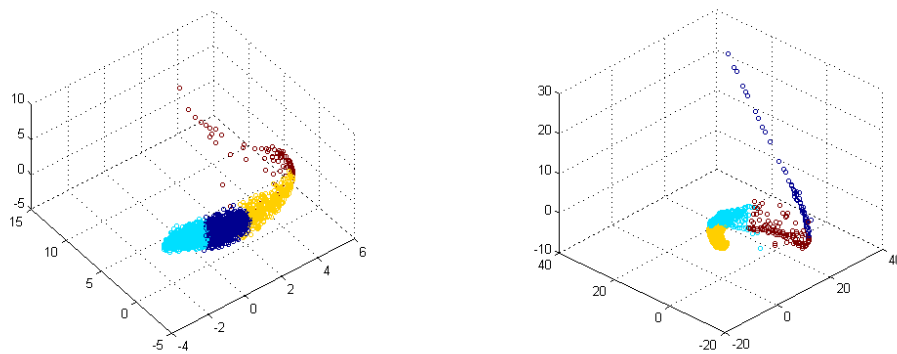


Figura 4.3: Sogliature di Fig. 4.1

Figura 4.4: Clustering per la funzione *distanza*Figura 4.5: Visualizzazione nello spazio bidimensionale per la funzione *distanza*, il colore indica il cluster di appartenenzaFigura 4.6: Visualizzazione nello spazio tridimensionale per la funzione *distanza*, il colore indica il cluster di appartenenza

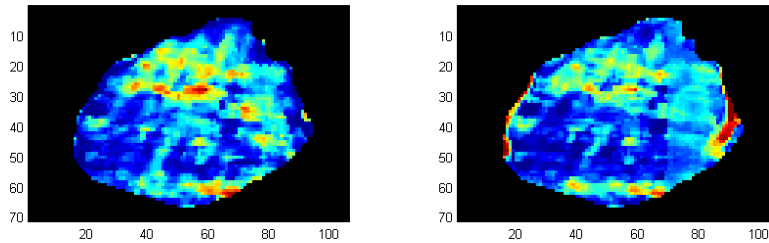


Figura 4.7: Esito della riduzione monodimensionale della funzione *peso*, a sinistra con allineamento *Best*, a destra *Manual* e riduzione con criterio di Sammon

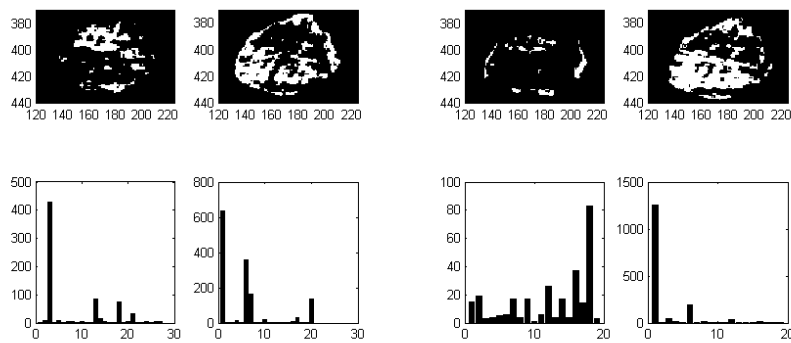


Figura 4.8: Sogliature di Fig. 4.1

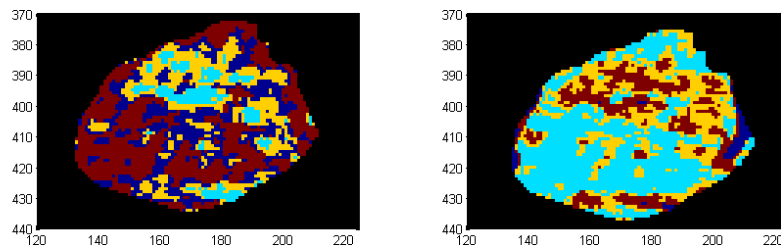


Figura 4.9: Clustering per la funzione *peso*

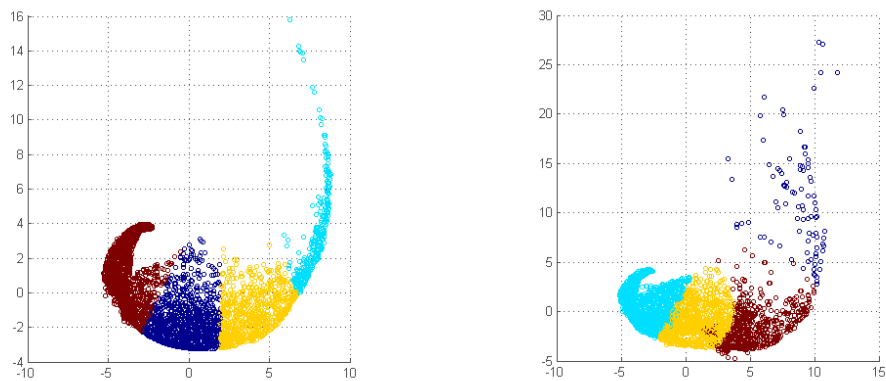


Figura 4.10: Visualizzazione nello spazio bidimensionale per la funzione *peso*, il colore indica il cluster di appartenenza

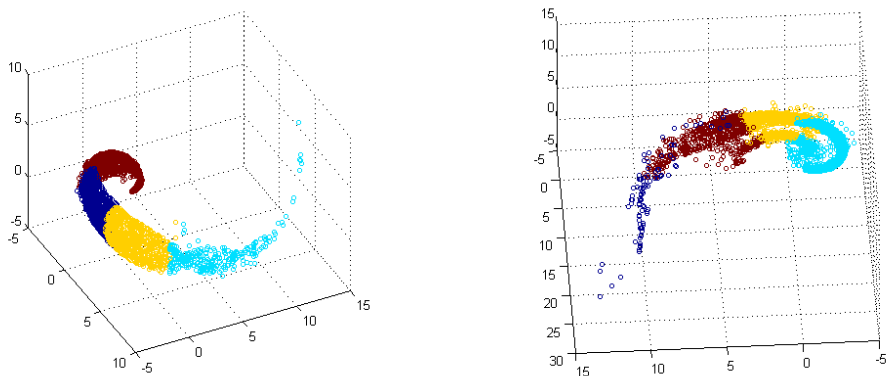


Figura 4.11: Visualizzazione nello spazio tridimensionale per la funzione *peso*, il colore indica il cluster di appartenenza

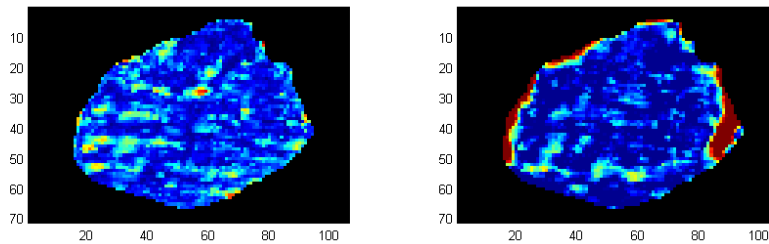


Figura 4.12: Esito della riduzione monodimensionale della funzione *varia*, a sinistra con allineamento *Best*, a destra *Manual* e riduzione con criterio di Sammon

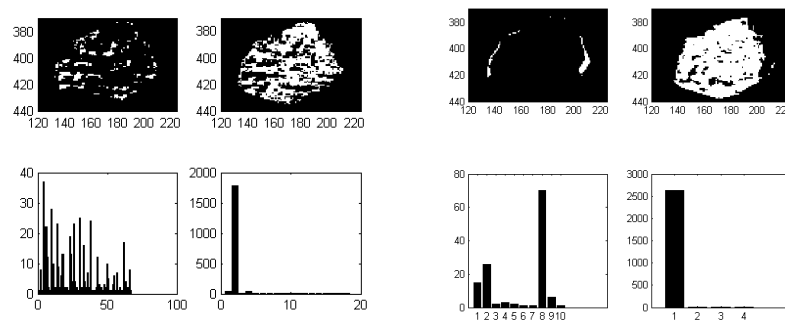
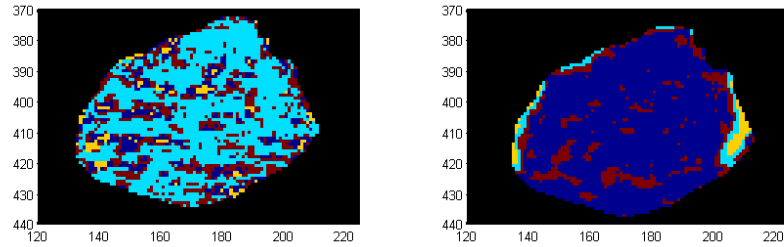
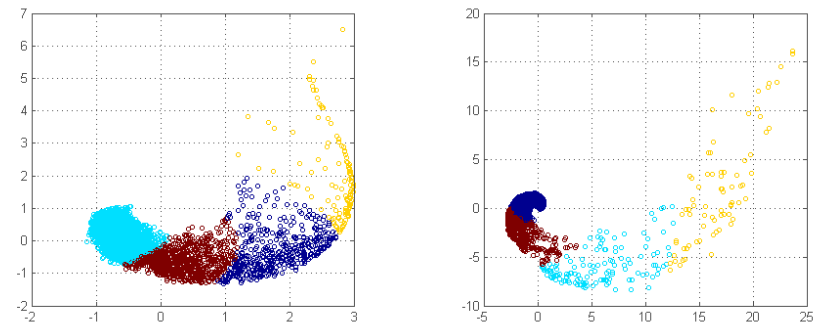
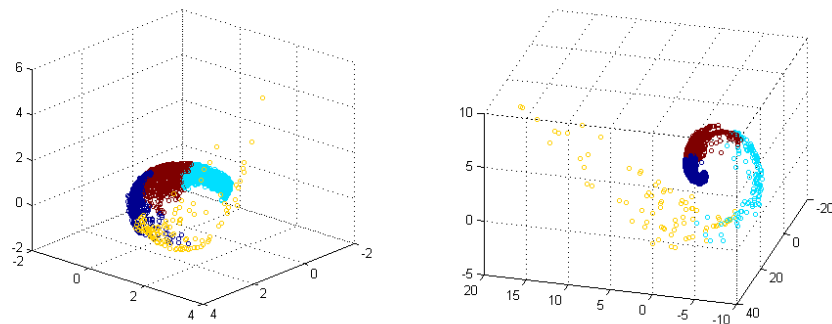


Figura 4.13: Sogliature di Fig. 4.1

Figura 4.14: Clustering per la funzione *varia*Figura 4.15: Visualizzazione sullo spazio bidimensionale per la funzione *varia*, il colore indica il cluster di appartenenzaFigura 4.16: Visualizzazione sullo spazio tridimensionale per la funzione *varia*, il colore indica il cluster di appartenenza

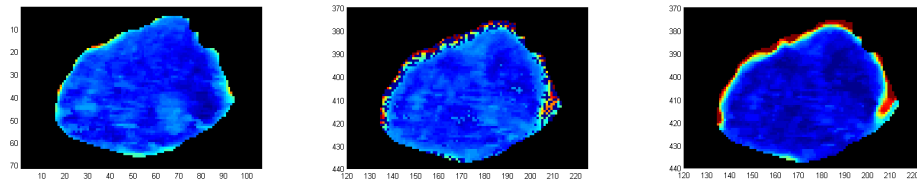


Figura 4.17: Esito della riduzione monodimensionale della funzione *altezza*, a sinistra con allineamento *Best*, a destra *Manual* e riduzione con criterio di Sammon

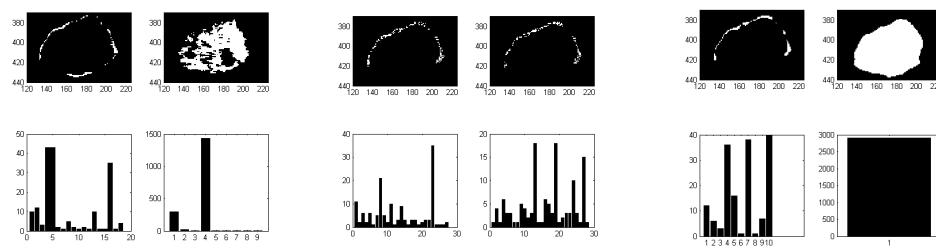


Figura 4.18: Sogliature di Fig. 4.1

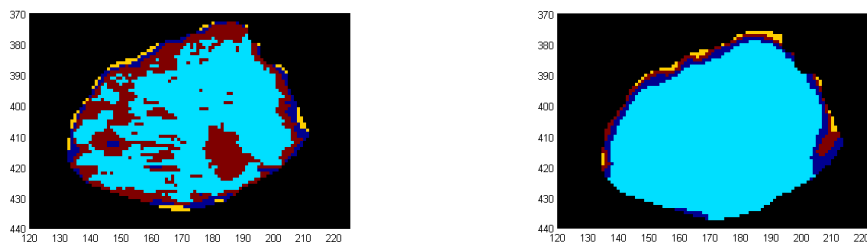


Figura 4.19: Clustering per la funzione *altezza*

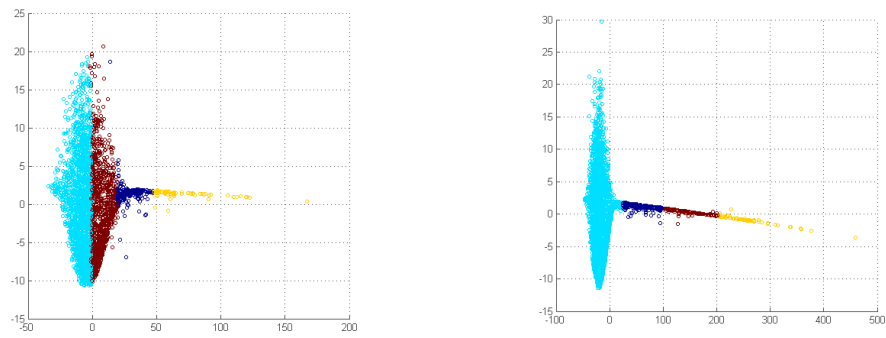


Figura 4.20: Visualizzazione sullo spazio bidimensionale per la funzione *altezza*, il colore indica il cluster di appartenenza

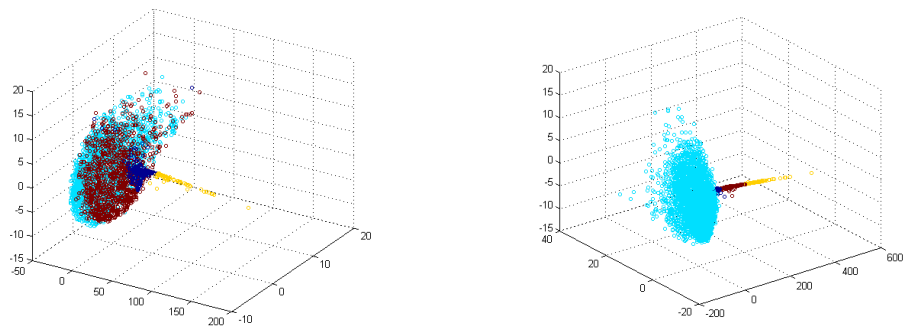


Figura 4.21: Visualizzazione sullo spazio tridimensionale per la funzione *altezza*, il colore indica il cluster di appartenenza



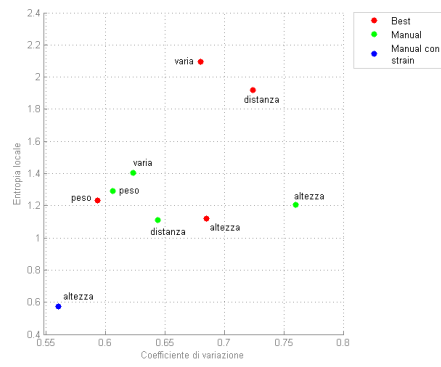


Figura 4.22: Grafico rappresentante gli indicatori di rumore per ciascuna analisi effettuata sul primo esame

## 4.2 Esame 2

Il secondo esame riguarda immagini TC perfusionali di un paziente affetto da tumore nel polmone. È costituito da 20 slice temporali e la ROI è formata da 4405 pixels.

Anche in questo caso le immagini dell'allineamento manuale presentano artefatti collocati sul bordo superiore e laterale sinistro. Anche in questo caso la presenza di tali pixels anomali compromette lievemente anche la zona più interna.

La particolarità di questo esame è la presenza di un'area, costituita da due componenti, come visibile nell'immagine corrispondente alla funzione *distanza* con allineamento.

Questa macchia corrisponde a un fenomeno visibile anche ad occhio nudo nelle prime slices dell'esame originale, come da 4.2. Ciò permette di affermare che il metodo qui utilizzato sia in grado, per lo meno parzialmente, di individuare comportamenti di perfusione non ordinari.

In questo esame è inoltre visibile come il criterio *strain* abbia un comportamento migliore nei casi di unificazione dei punti. Mantenendo il criterio di Sammon compare una granulosità non compatibile con la clusterizzazione, che è effettuata a priori dell'applicazione dei criteri.

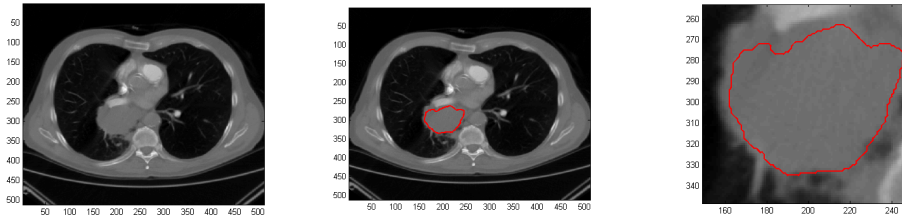


Figura 4.23: Da sinistra a destra un fotogramma completo, la maschera all'interno di esso e un ingrandimento della ROI

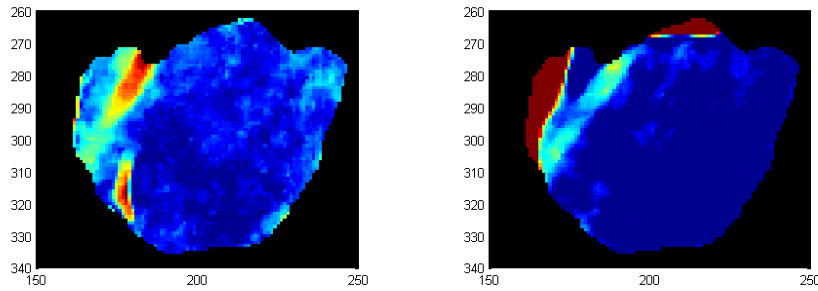


Figura 4.24: Esito della riduzione monodimensionale della funzione *distanza*, a sinistra con allineamento *Best*, a destra *Manual* e riduzione con criterio di Sammon

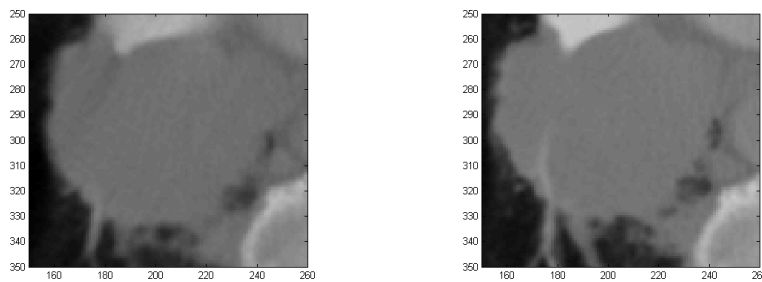


Figura 4.25: Prime due slice del secondo esame. Si può notare in alto a sinistra nella prima e in basso a sinistra nella seconda le due zone di intensità di grigio ben differenti dal resto della maschera

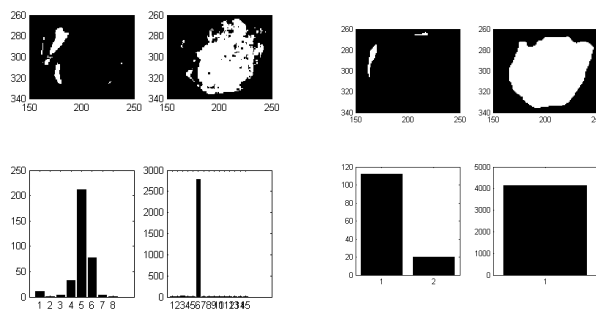


Figura 4.26: Sogliature della prima riga di Fig. 4.2

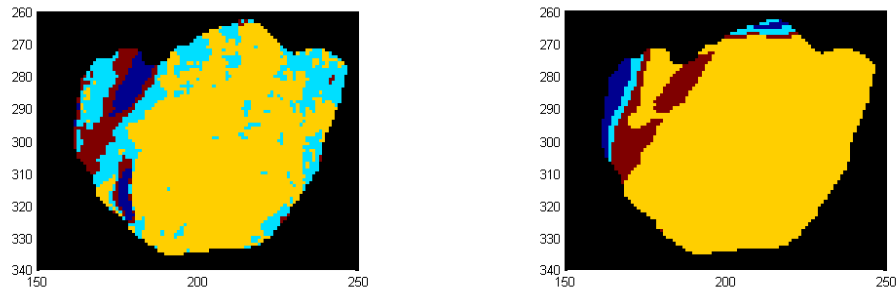


Figura 4.27: Clustering per la funzione *distanza*

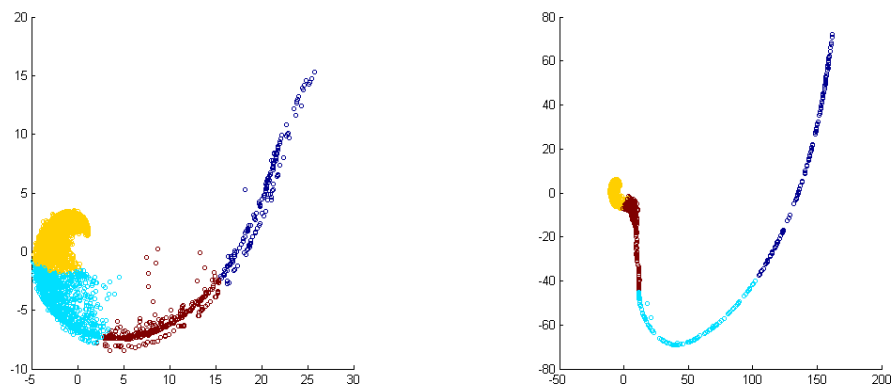


Figura 4.28: Visualizzazione sullo spazio bidimensionale per la funzione *distanza*, il colore indica il cluster di appartenenza

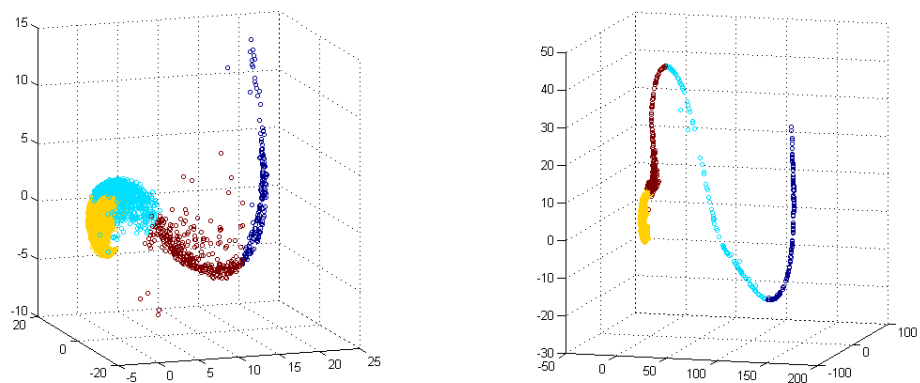


Figura 4.29: Visualizzazione sullo spazio tridimensionale per la funzione *distanza*, il colore indica il cluster di appartenenza

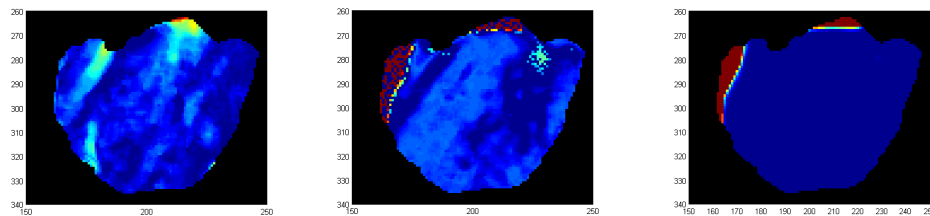


Figura 4.30: Esito della riduzione monodimensionale della funzione *peso*, a sinistra con allineamento *Best*, al centro *Manual* e riduzione con criterio di Sammon e a destra *Manual* e riduzione con criterio strain

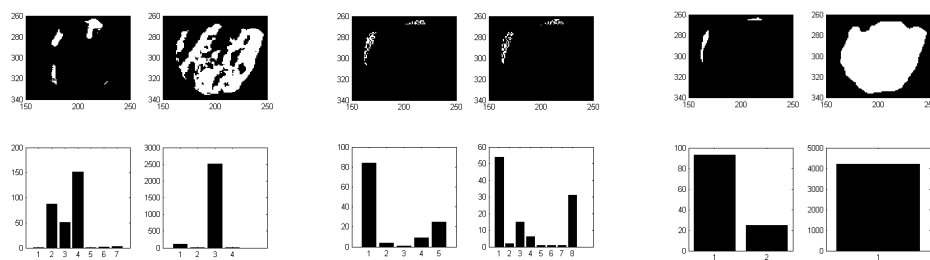


Figura 4.31: Sogliature di Fig. 4.2

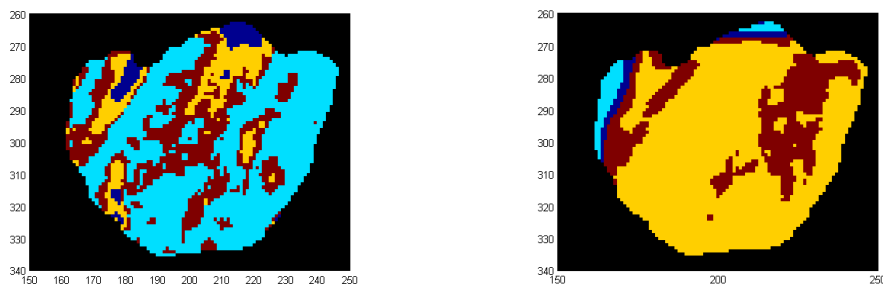


Figura 4.32: Clustering per la funzione *peso*

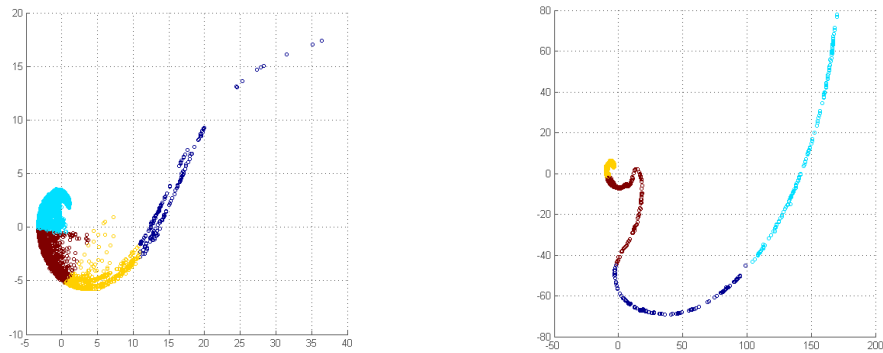


Figura 4.33: Visualizzazione sullo spazio bidimensionale per la funzione *peso*, il colore indica il cluster di appartenenza

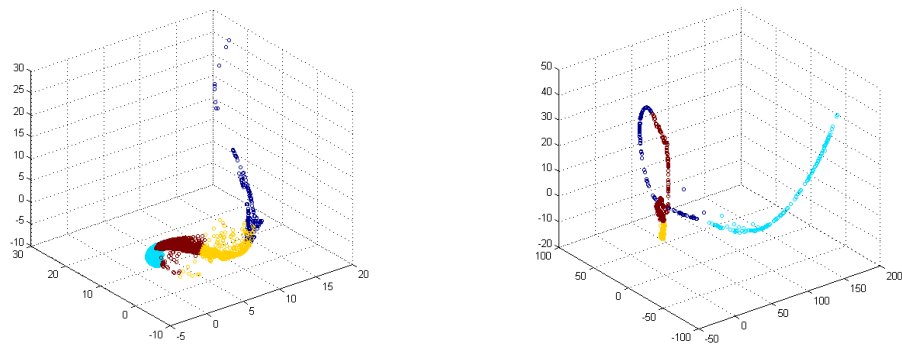


Figura 4.34: Visualizzazione sullo spazio tridimensionale per la funzione *peso*, il colore indica il cluster di appartenenza

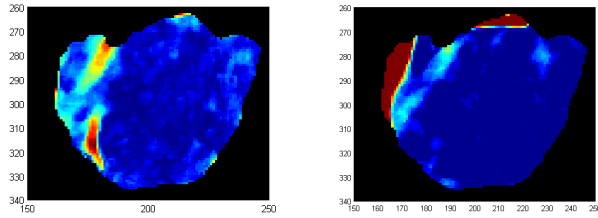


Figura 4.35: Esito della riduzione monodimensionale della funzione *varia*, a sinistra con allineamento *Best*, a destra *Manual* e riduzione con criterio di Sammon

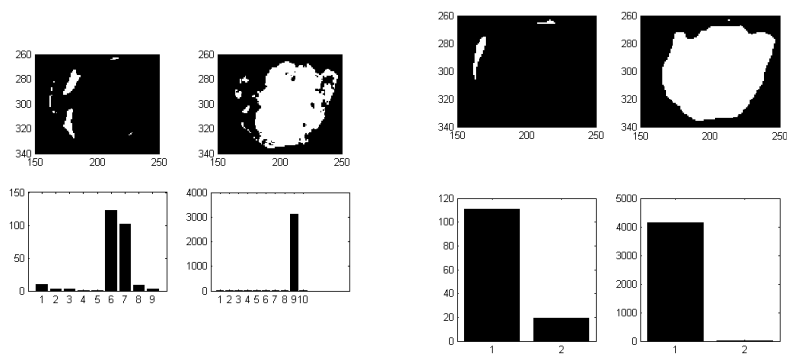


Figura 4.36: Sogliature di Fig. 4.2

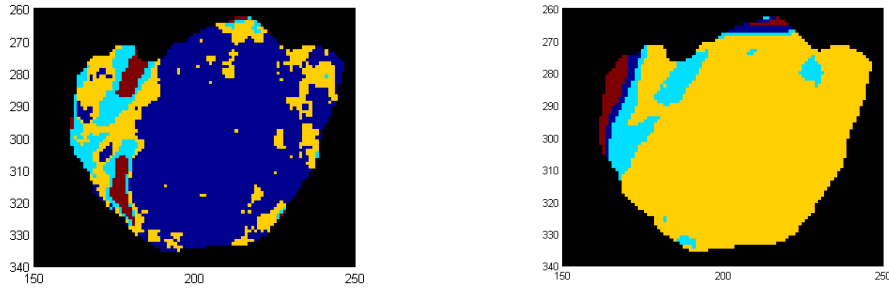


Figura 4.37: Clustering per la funzione *varia*

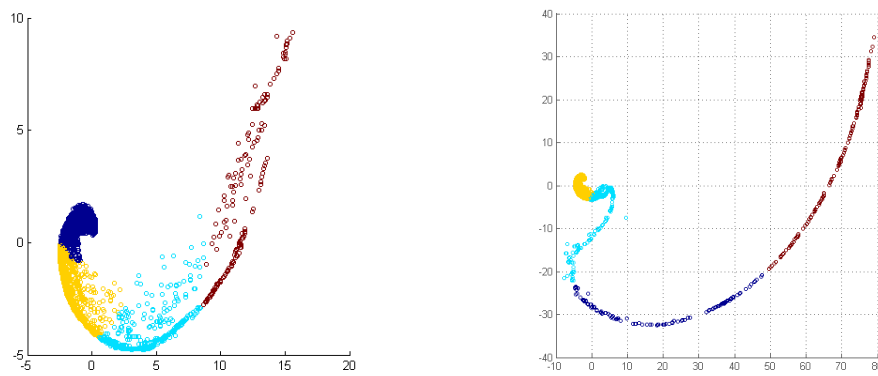


Figura 4.38: Visualizzazione sullo spazio bidimensionale per la funzione *varia*, il colore indica il cluster di appartenenza

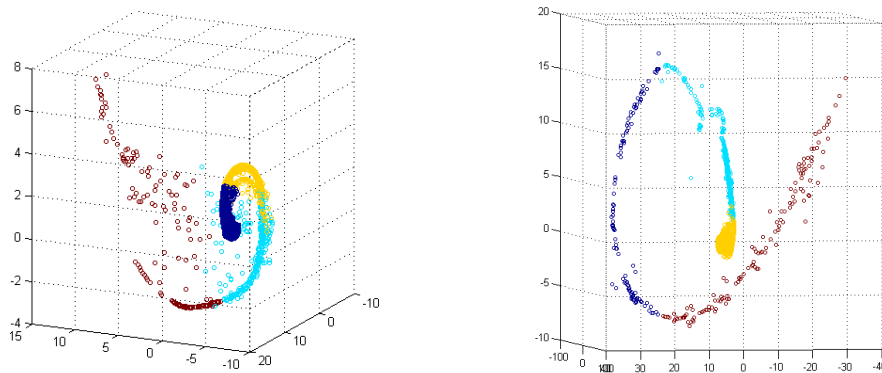


Figura 4.39: Visualizzazione sullo spazio tridimensionale per la funzione *varia*, il colore indica il cluster di appartenenza



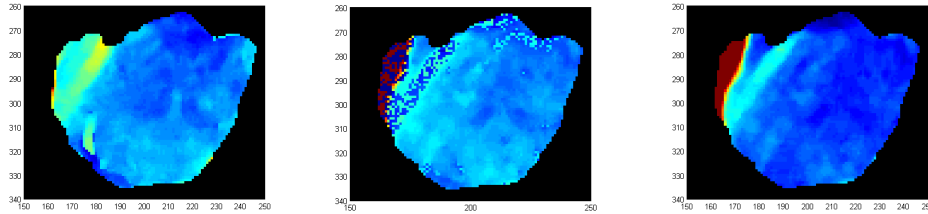


Figura 4.40: Esito della riduzione monodimensionale della funzione *altezza*, a sinistra con allineamento *Best*, al centro *Manual* e riduzione con criterio di Sammon e a destra *Manual* e riduzione con criterio strain

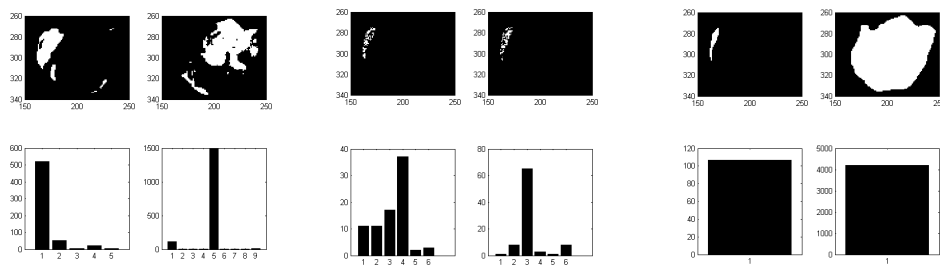


Figura 4.41: Sogliature di Fig. 4.2

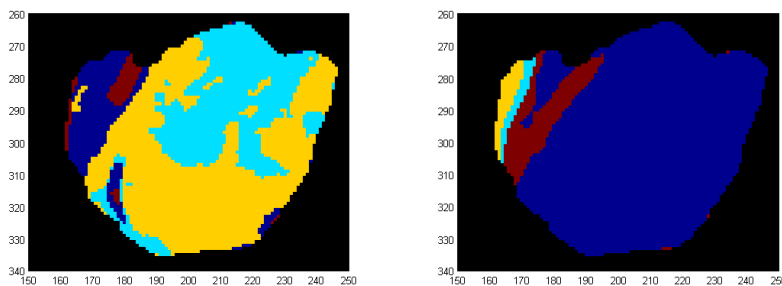


Figura 4.42: Clustering per la funzione *altezza*

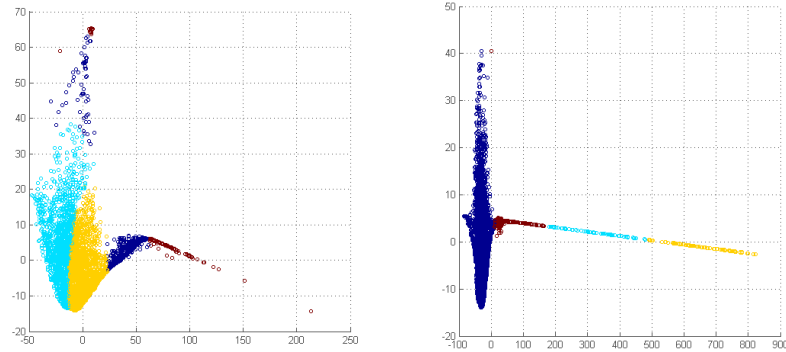


Figura 4.43: Visualizzazione sullo spazio bidimensionale per la funzione *altezza*, il colore indica il cluster di appartenenza

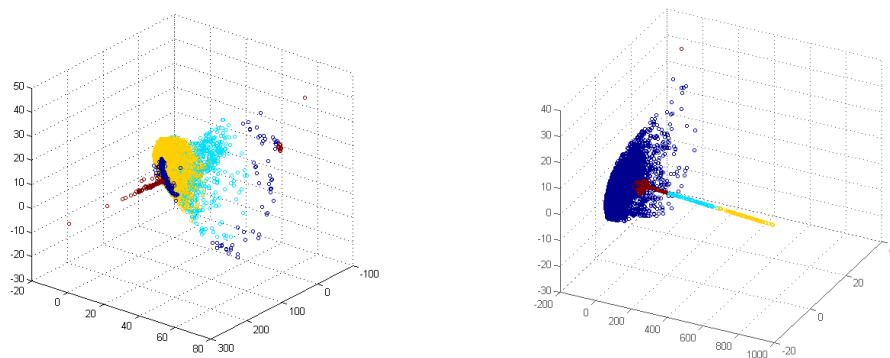


Figura 4.44: Visualizzazione sullo spazio tridimensionale per la funzione *altezza*, il colore indica il cluster di appartenenza

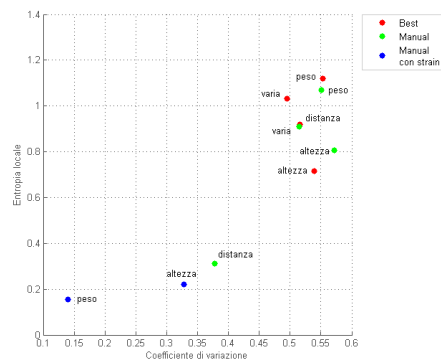


Figura 4.45: Grafico rappresentante gli indicatori di rumore per ciascuna analisi effettuata sul secondo esame

### 4.3 Esame 3

L'ultimo esame riguarda immagini TC perfusionali di un paziente affetto da tumore al fegato. È costituito da 20 slice temporali e la ROI è formata da 864 pixels.

In figura 4.3, 4.3,4.3,4.3 sono mostrate le immagini relative alla visualizzazione ottenuta con riduzione monodimensionale.

La visualizzazione che si distingue maggiormente è quella corrispondente alla funzione *distanza* con allineamento *Best*. In questa è molto accentuata la presenza di pixels disposti dal bordo con comportamento molto diverso rispetto alla zona centrale. Questo effetto è riscontrabile in modo meno accentuato anche su altre immagini

Per questo esame abbiamo anche un riscontro con le analisi effettuate dal CVG. In figura 4.3 è mostrata una colormap ottenuta con un modello sigmoideale: è ben visibile la zona in blu compatibile con la zona blu di figura 4.3.

Anche la funzione *peso* offre una distribuzione più regolare rispetto alle altre, mantenendo una correlazione con l'immagine 4.3. Nelle sezioni seguenti sono proposti gli altri metodi di visualizzazione.

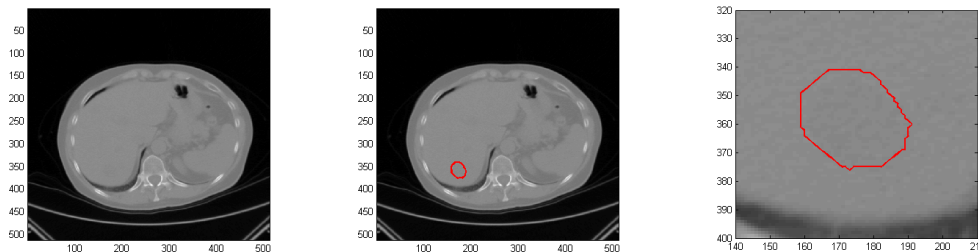


Figura 4.46: Da sinistra a destra un fotogramma completo, la maschera all'interno di esso e un ingrandimento della ROI

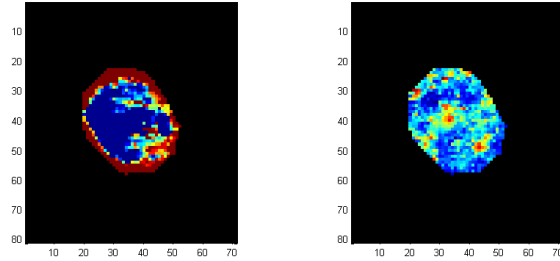


Figura 4.47: Esito della riduzione monodimensionale della funzione *distanza*, a sinistra con allineamento *Best*, a destra *Manual* e riduzione con criterio di Sammon

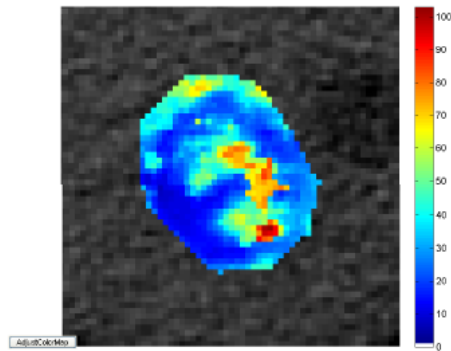


Figura 4.48: Colormap ottenuta dal CVG, con un modello sigmoidale

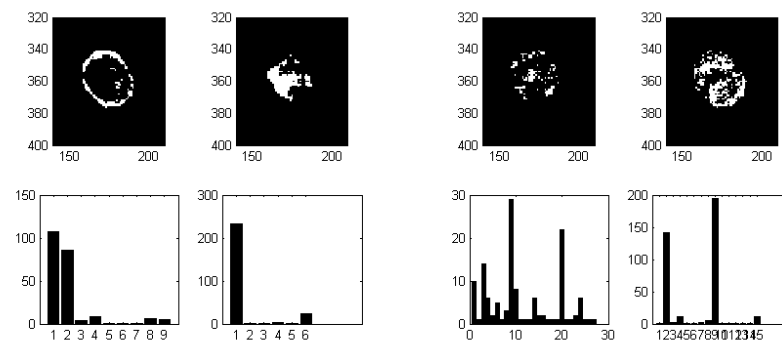
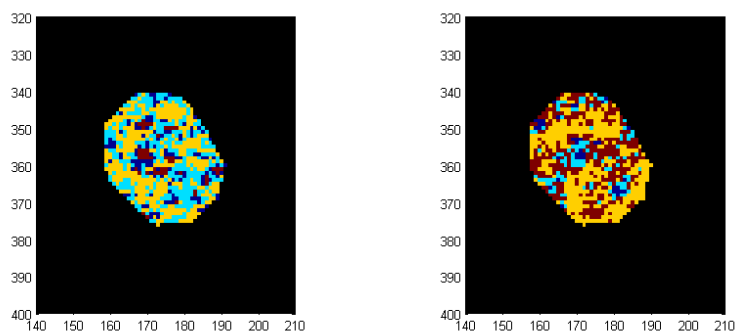
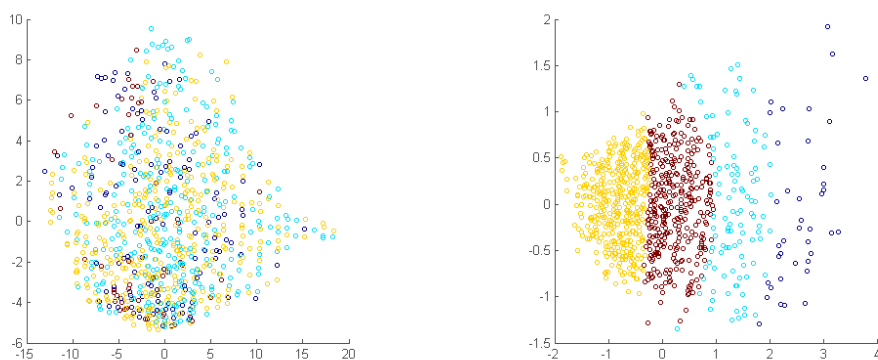
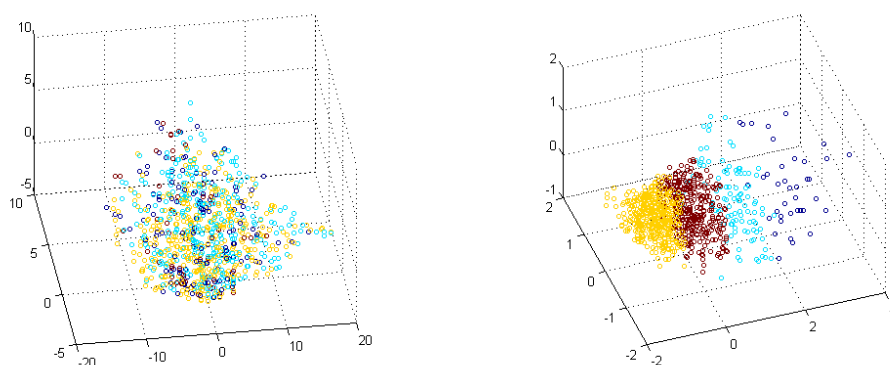


Figura 4.49: Sogliature di Fig. 4.3

Figura 4.50: Clustering per la funzione *distanza*Figura 4.51: Visualizzazione nello spazio bidimensionale per la funzione *distanza*, il colore indica il cluster di appartenenzaFigura 4.52: Visualizzazione nello spazio tridimensionale per la funzione *distanza*, il colore indica il cluster di appartenenza

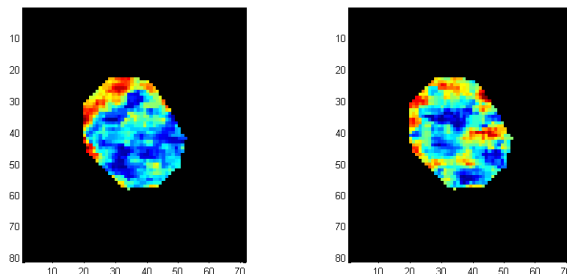


Figura 4.53: Esito della riduzione monodimensionale della funzione *peso*, a sinistra con allineamento *Best*, a destra *Manual* e riduzione con criterio di Sammon

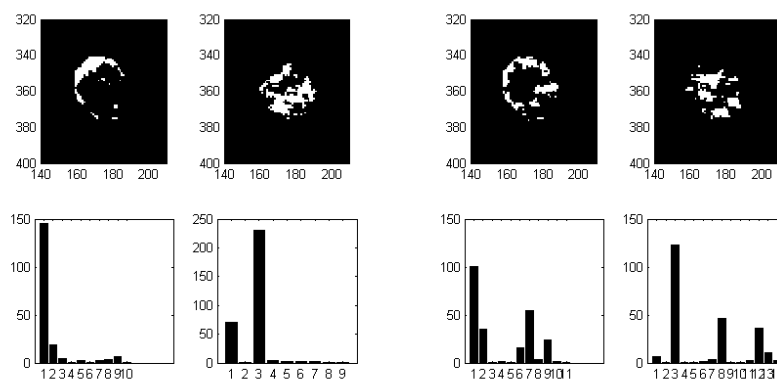


Figura 4.54: Sogliature di Fig. 4.3

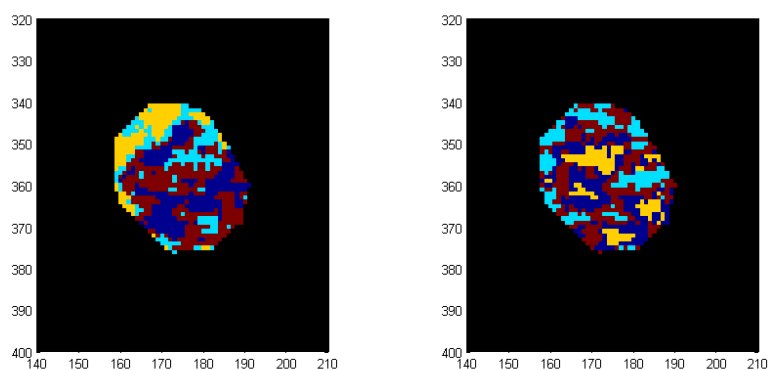


Figura 4.55: Clustering per la funzione *peso*

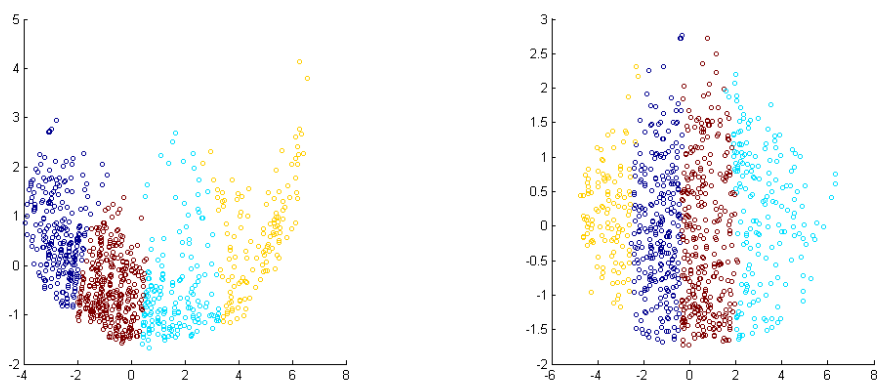


Figura 4.56: Visualizzazione nello spazio bidimensionale per la funzione *peso*, il colore indica il cluster di appartenenza

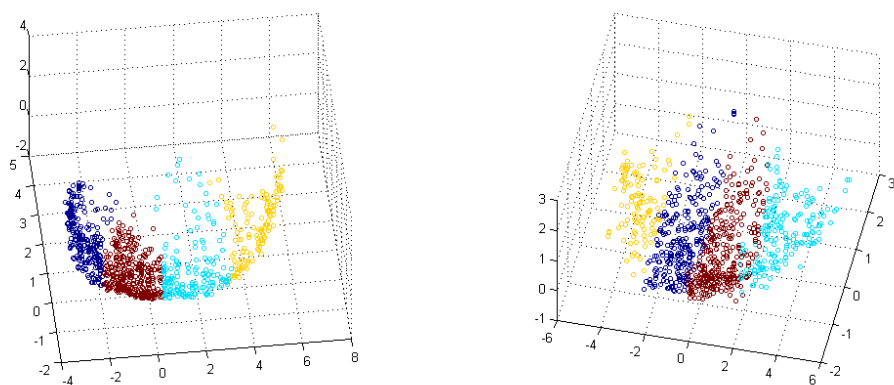


Figura 4.57: Visualizzazione nello spazio tridimensionale per la funzione *peso*, il colore indica il cluster di appartenenza

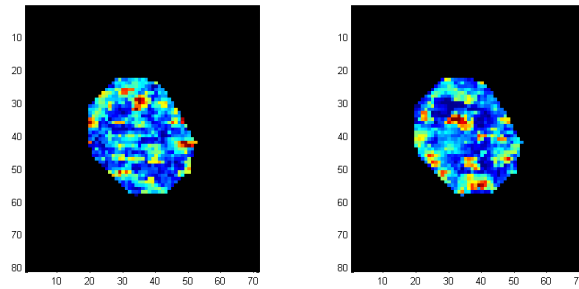


Figura 4.58: Esito della riduzione monodimensionale della funzione *varia*, a sinistra con allineamento *Best*, a destra *Manual* e riduzione con criterio di Sammon

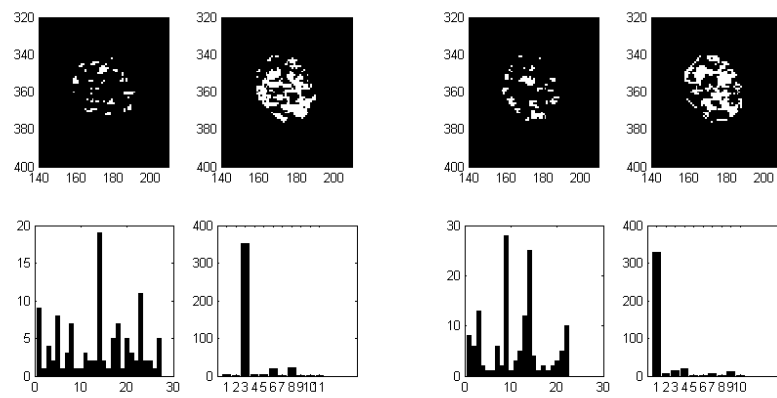
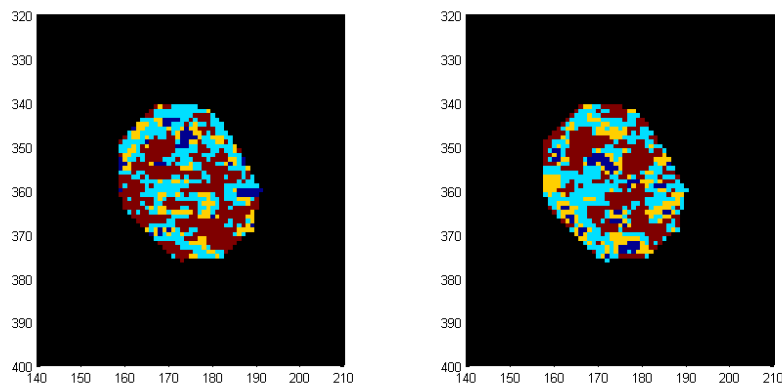
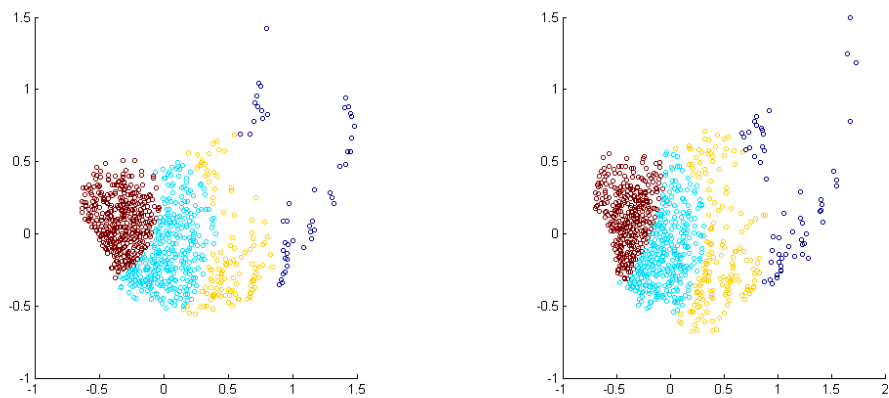
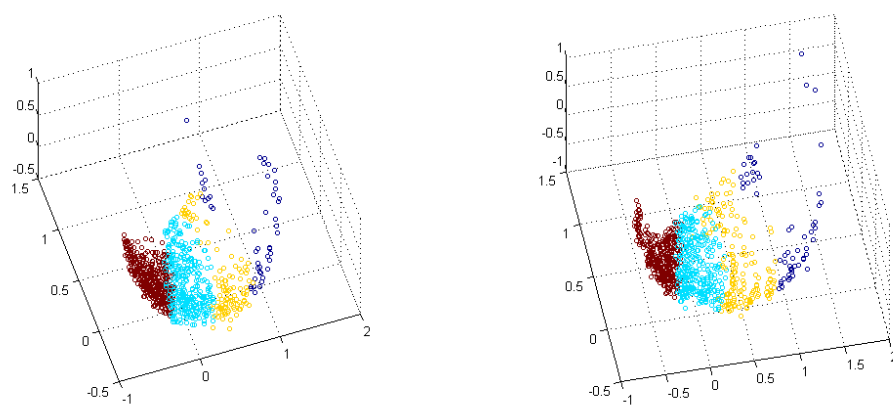


Figura 4.59: Sogliature di Fig. 4.3



Figura 4.60: Clustering per la funzione *varia*Figura 4.61: Visualizzazione sullo spazio bidimensionale per la funzione *varia*, il colore indica il cluster di appartenenzaFigura 4.62: Visualizzazione sullo spazio tridimensionale per la funzione *varia*, il colore indica il cluster di appartenenza

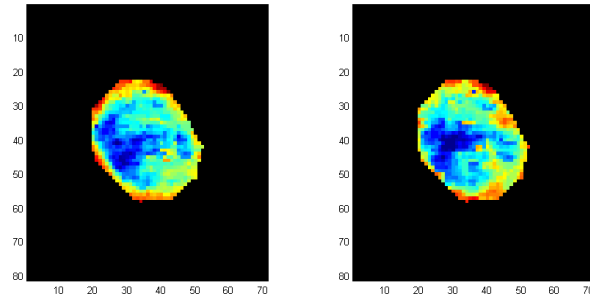


Figura 4.63: Esito della riduzione monodimensionale della funzione *altezza*, a sinistra con allineamento *Best*, a destra *Manual* e riduzione con criterio di Sammon

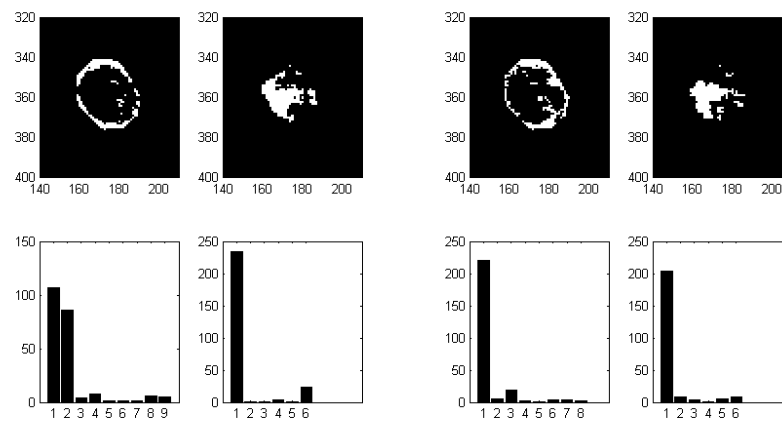


Figura 4.64: Sogliature di Fig. 4.3

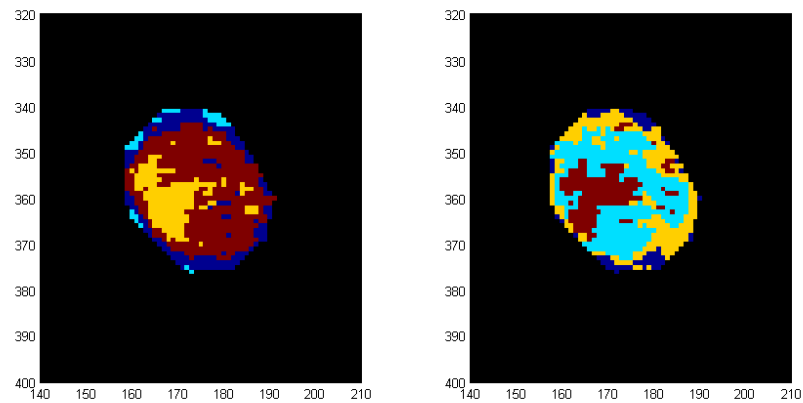


Figura 4.65: Clustering per la funzione *altezza*

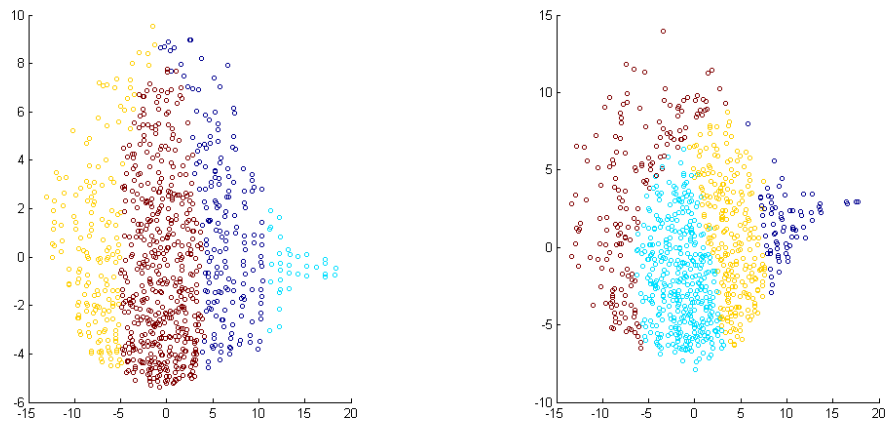


Figura 4.66: Visualizzazione sullo spazio bidimensionale per la funzione *altezza*, il colore indica il cluster di appartenenza

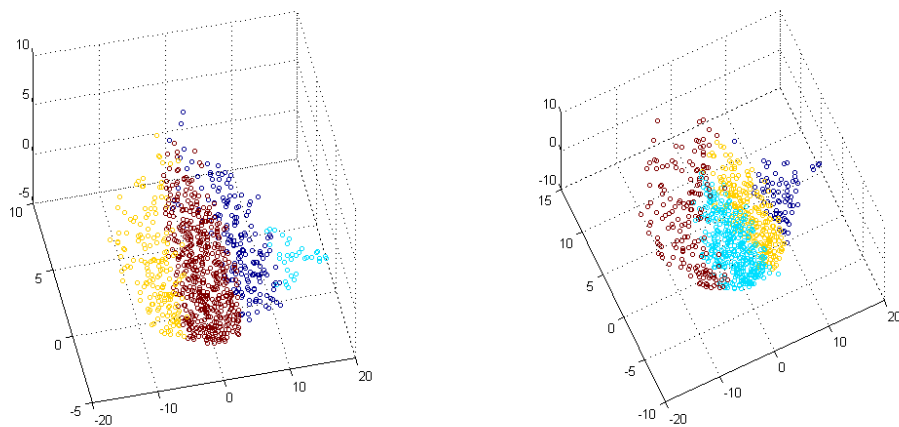


Figura 4.67: Visualizzazione sullo spazio tridimensionale per la funzione *altezza*, il colore indica il cluster di appartenenza

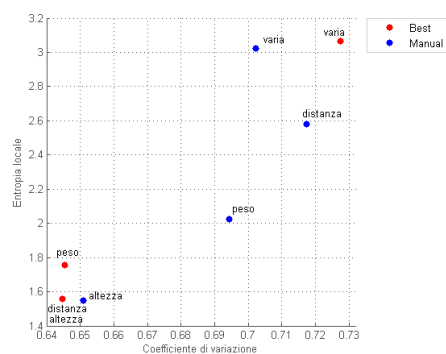


Figura 4.68: Grafico rappresentante gli indicatori di rumore per ciascuna analisi effettuata sul terzo esame

# Conclusioni

L'analisi sviluppata in questa tesi propone una possibilità di inagine delle immagini perfusionali differente da quella già presente in letteratura. I risultati ottenuti suggeriscono la presenza di informazione rilevante, per lo meno in parte, per compatibilità con le analisi precedenti e con le radiografie di partenza. Tuttavia, in quanto analisi preliminare, dista molto dall'essere un risultato netto e necessita di numerosi miglioramenti: il costo computazionale totale dell'elaborazione è ancora molto oneroso a causa dei numerosi confronti singoli tra pixels. Inoltre sono da rivalutare le analisi di qualità dei risultati ottenuti.

# Indice

<b>Introduzione</b>	<b>i</b>
<b>1 Premesse</b>	<b>1</b>
1.1 Tumori . . . . .	1
1.2 Diagnosi . . . . .	2
1.3 Acquisizione dati e modelli . . . . .	4
<b>2 Impianto teorico-matematico</b>	<b>7</b>
2.1 Teoria della funzione di taglia . . . . .	9
2.2 Riduzione dimensionale . . . . .	13
2.3 Clustering . . . . .	14
<b>3 Svolgimento dell'elaborazione</b>	<b>15</b>
3.1 Scelta dello spazio topologico . . . . .	15
3.2 Scelte di funzione misurante . . . . .	17
3.3 Confronto . . . . .	17
3.4 Visualizzazione . . . . .	17
3.5 Analisi di qualità . . . . .	18
<b>4 Esiti delle sperimentazioni</b>	<b>21</b>
4.1 Esame 1 . . . . .	22
4.2 Esame 2 . . . . .	32
4.3 Esame 3 . . . . .	41
<b>Bibliografia</b>	<b>7</b>



# Elenco delle figure

3.1	Intensità di perfusione di un singolo pixel . . . . .	16
4.1	Da sinistra a destra un fotogramma completo, la maschera all'interno di esso e un ingrandimento della ROI . . . . .	23
4.2	Esito della riduzione monodimensionale della funzione <i>distanza</i> , a sinistra con allineamento <i>Best</i> , a destra <i>Manual</i> e riduzione con criterio di Sammon . . . . .	23
4.3	Sogliature di Fig. 4.1 . . . . .	23
4.4	Clustering per la funzione <i>distanza</i> . . . . .	24
4.5	Visualizzazione nello spazio bidimensionale per la funzione <i>distanza</i> , il colore indica il cluster di appartenenza . . . . .	24
4.6	Visualizzazione nello spazio tridimensionale per la funzione <i>distanza</i> , il colore indica il cluster di appartenenza . . . . .	24
4.7	Esito della riduzione monodimensionale della funzione <i>peso</i> , a sinistra con allineamento <i>Best</i> , a destra <i>Manual</i> e riduzione con criterio di Sammon . . . . .	25
4.8	Sogliature di Fig. 4.1 . . . . .	25
4.9	Clustering per la funzione <i>peso</i> . . . . .	25
4.10	Visualizzazione nello spazio bidimensionale per la funzione <i>peso</i> , il colore indica il cluster di appartenenza . . . . .	26
4.11	Visualizzazione nello spazio tridimensionale per la funzione <i>peso</i> , il colore indica il cluster di appartenenza . . . . .	26
4.12	Esito della riduzione monodimensionale della funzione <i>varia</i> , a sinistra con allineamento <i>Best</i> , a destra <i>Manual</i> e riduzione con criterio di Sammon . . . . .	27
4.13	Sogliature di Fig. 4.1 . . . . .	27
4.14	Clustering per la funzione <i>varia</i> . . . . .	28



4.15	Visualizzazione sullo spazio bidimensionale per la funzione <i>varia</i> , il colore indica il cluster di appartenenza . . . . .	28
4.16	Visualizzazione sullo spazio tridimensionale per la funzione <i>varia</i> , il colore indica il cluster di appartenenza . . . . .	28
4.17	Esito della riduzione monodimensionale della funzione <i>altezza</i> , a sinistra con allineamento <i>Best</i> , a destra <i>Manual</i> e riduzione con criterio di Sammon . . . . .	29
4.18	Sogliature di Fig. 4.1 . . . . .	29
4.19	Clustering per la funzione <i>altezza</i> . . . . .	29
4.20	Visualizzazione sullo spazio bidimensionale per la funzione <i>altezza</i> , il colore indica il cluster di appartenenza . . . . .	30
4.21	Visualizzazione sullo spazio tridimensionale per la funzione <i>altezza</i> , il colore indica il cluster di appartenenza . . . . .	30
4.22	Grafico rappresentante gli indicatori di rumore per ciascuna analisi effettuata sul primo esame . . . . .	31
4.23	Da sinistra a destra un fotogramma completo, la maschera all'interno di esso e un ingrandimento della ROI . . . . .	32
4.24	Esito della riduzione monodimensionale della funzione <i>distanza</i> , a sinistra con allineamento <i>Best</i> , a destra <i>Manual</i> e riduzione con criterio di Sammon . . . . .	33
4.25	Prime due slice del secondo esame. Si può notare in alto a sinistra nella prima e in basso a sinistra nella seconda le due zone di intensità di grigio ben differnti dal resto della maschera	33
4.26	Sogliature della prima riga di Fig. 4.2 . . . . .	33
4.27	Clustering per la funzione <i>distanza</i> . . . . .	34
4.28	Visualizzazione sullo spazio bidimensionale per la funzione <i>distanza</i> , il colore indica il cluster di appartenenza . . . . .	34
4.29	Visualizzazione sullo spazio tridimensionale per la funzione <i>distanza</i> , il colore indica il cluster di appartenenza . . . . .	34
4.30	Esito della riduzione monodimensionale della funzione <i>peso</i> , a sinistra con allineamento <i>Best</i> , al centro <i>Manual</i> e riduzione con criterio di Sammon e a destra <i>Manual</i> e riduzione con criterio strain . . . . .	35
4.31	Sogliature di Fig. 4.2 . . . . .	35
4.32	Clustering per la funzione <i>peso</i> . . . . .	35
4.33	Visualizzazione sullo spazio bidimensionale per la funzione <i>peso</i> , il colore indica il cluster di appartenenza . . . . .	36

---

4.34	Visualizzazione sullo spazio tridimensionale per la funzione <i>peso</i> , il colore indica il cluster di appartenenza . . . . .	36
4.35	Esito della riduzione monodimensionale della funzione <i>varia</i> , a sinistra con allineamento <i>Best</i> , a destra <i>Manual</i> e riduzione con criterio di Sammon . . . . .	37
4.36	Sogliature di Fig. 4.2 . . . . .	37
4.37	Clustering per la funzione <i>varia</i> . . . . .	38
4.38	Visualizzazione sullo spazio bidimensionale per la funzione <i>varia</i> , il colore indica il cluster di appartenenza . . . . .	38
4.39	Visualizzazione sullo spazio tridimensionale per la funzione <i>varia</i> , il colore indica il cluster di appartenenza . . . . .	38
4.40	Esito della riduzione monodimensionale della funzione <i>altezza</i> , a sinistra con allineamento <i>Best</i> , al centro <i>Manual</i> e riduzione con criterio di Sammon e a destra <i>Manual</i> e riduzione con criterio strain . . . . .	39
4.41	Sogliature di Fig. 4.2 . . . . .	39
4.42	Clustering per la funzione <i>altezza</i> . . . . .	39
4.43	Visualizzazione sullo spazio bidimensionale per la funzione <i>altezza</i> , il colore indica il cluster di appartenenza . . . . .	40
4.44	Visualizzazione sullo spazio tridimensionale per la funzione <i>altezza</i> , il colore indica il cluster di appartenenza . . . . .	40
4.45	Grafico rappresentante gli indicatori di rumore per ciascuna analisi effettuata sul secondo esame . . . . .	40
4.46	Da sinistra a destra un fotogramma completo, la maschera all'interno di esso e un ingrandimento della ROI . . . . .	41
4.47	Esito della riduzione monodimensionale della funzione <i>distanza</i> , a sinistra con allineamento <i>Best</i> , a destra <i>Manual</i> e riduzione con criterio di Sammon . . . . .	42
4.48	Colormap ottenuta dal CVG, con un modello sigmoidale . . . . .	42
4.49	Sogliature di Fig. 4.3 . . . . .	42
4.50	Clustering per la funzione <i>distanza</i> . . . . .	43
4.51	Visualizzazione nello spazio bidimensionale per la funzione <i>distanza</i> , il colore indica il cluster di appartenenza . . . . .	43
4.52	Visualizzazione nello spazio tridimensionale per la funzione <i>distanza</i> , il colore indica il cluster di appartenenza . . . . .	43

4.53	Esito della riduzione monodimensionale della funzione <i>peso</i> , a sinistra con allineamento <i>Best</i> , a destra <i>Manual</i> e riduzione con criterio di Sammon . . . . .	44
4.54	Sogliature di Fig. 4.3 . . . . .	44
4.55	Clustering per la funzione <i>peso</i> . . . . .	44
4.56	Visualizzazione nello spazio bidimensionale per la funzione <i>peso</i> , il colore indica il cluster di appartenenza . . . . .	45
4.57	Visualizzazione nello spazio tridimensionale per la funzione <i>peso</i> , il colore indica il cluster di appartenenza . . . . .	45
4.58	Esito della riduzione monodimensionale della funzione <i>varia</i> , a sinistra con allineamento <i>Best</i> , a destra <i>Manual</i> e riduzione con criterio di Sammon . . . . .	46
4.59	Sogliature di Fig. 4.3 . . . . .	46
4.60	Clustering per la funzione <i>varia</i> . . . . .	47
4.61	Visualizzazione sullo spazio bidimensionale per la funzione <i>varia</i> , il colore indica il cluster di appartenenza . . . . .	47
4.62	Visualizzazione sullo spazio tridimensionale per la funzione <i>varia</i> , il colore indica il cluster di appartenenza . . . . .	47
4.63	Esito della riduzione monodimensionale della funzione <i>altezza</i> , a sinistra con allineamento <i>Best</i> , a destra <i>Manual</i> e riduzione con criterio di Sammon . . . . .	48
4.64	Sogliature di Fig. 4.3 . . . . .	48
4.65	Clustering per la funzione <i>altezza</i> . . . . .	48
4.66	Visualizzazione sullo spazio bidimensionale per la funzione <i>altezza</i> , il colore indica il cluster di appartenenza . . . . .	49
4.67	Visualizzazione sullo spazio tridimensionale per la funzione <i>altezza</i> , il colore indica il cluster di appartenenza . . . . .	49
4.68	Grafico rappresentante gli indicatori di rumore per ciascuna analisi effettuata sul terzo esame . . . . .	50

# Bibliografia

- [1] Frosini P., *A distance for similarity classes of submanifolds of a euclidean space*, Bull. Austral. Math. Soc., Vol. 42, (1990), 407-416,
- [2] Verri A., Uras C., Frosini P., Ferri M., *On the use of size functions for shape analysis*, Biological Cybernetics, 70, (1993), 99-107.
- [3] Frosini P. , Landi C., *Size functions and formal series*, Applicable Algebra in Engineering, Communication and Computing, 12(4) (2001), 327-349.
- [4] Frosini P., Landi C., *Size theory as a topological tool for computer vision*, Pattern Recognition And Image Analysis, vol. 9 (4) (1999), 596-603.
- [5] D'Amico M., Frosini P., Landi C., *Natural pseudo-distance and optimal matching between reduced size functions*, Acta Applicandae Mathematicae, vol. 109 (2010), n. 2, 527-554.
- [6] Ferri M., Lombardini S., Pallotti C., *Leuckocyte classification by size functions*, Proceedings of the seond IEEE Workshop on Applications of Computer Vision, IEEE Computer Society Press, Los Alamitos, CA (1994), 223-229
- [7] Uras C., Verri A., *On the recognition of the alphabet of the sign language through size functions*, Proc. XII IAPR International Conference of Pattern Recognition, Jerusalem (Israel). II, IEEE Computer Society Press, Los Alamitos, CA (1994), 334-338
- [8] Goutelle S., et al. *The Hill equation: a review of its capabilities in pharmacological modelling.*, Fundamental & clinical pharmacology 22.6 (2008), 633-648.

- [9] Jain, Anil K., and Richard C. Dubes. *Algorithms for clustering data*. Prentice-Hall, Inc., 1988.
- [10] MacQueen, J. *Some methods for classification and analysis of multivariate observations*. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, 281–297, University of California Press, Berkeley, Calif., 1967.