



*Consiglio Nazionale delle Ricerche
Istituto di Calcolo e Reti ad Alte Prestazioni*

Un algoritmo per l'interpolazione e la fusione multimodale di immagini medicali tridimensionali

I. Marra – M.Ciampi – A.Minutolo

RT-ICAR-NA-08-01

gennaio 2008



Consiglio Nazionale delle Ricerche, Istituto di Calcolo e Reti ad Alte Prestazioni (ICAR)
– Sede di Napoli, Via P. Castellino 111, 80131 Napoli, URL: www.na.icar.cnr.it



*Consiglio Nazionale delle Ricerche
Istituto di Calcolo e Reti ad Alte Prestazioni*

Un algoritmo per l'interpolazione e la fusione multimodale di immagini medicali tridimensionali

I. Marra¹ – M.Ciampi¹ – A.Minutolo¹

Rapporto Tecnico N.:
RT-ICAR-NA-08-01

Data:
gennaio 2008

¹ Istituto di Calcolo e Reti ad Alte Prestazioni, ICAR-CNR, Sede di Napoli, Via P. Castellino 111, 80131 Napoli

I rapporti tecnici dell'ICAR-CNR sono pubblicati dall'Istituto di Calcolo e Reti ad Alte Prestazioni del Consiglio Nazionale delle Ricerche. Tali rapporti, approntati sotto l'esclusiva responsabilità scientifica degli autori, descrivono attività di ricerca del personale e dei collaboratori dell'ICAR, in alcuni casi in un formato preliminare prima della pubblicazione definitiva in altra sede.

Un algoritmo per l'interpolazione e la fusione multimodale di immagini medicali tridimensionali

Mario Ciampi¹, Ivana Marra¹, Aniello Minutolo¹

¹ICAR-CNR, Via P. Castellino 111, 80131 Napoli, Italia
{mario.ciampi, ivana.marra, aniello.minutolo}@na.icar.cnr.it

Abstract

Una serie di immagini biomedicali (slices) prodotte da un esame clinico può essere vista come una singola immagine tridimensionale della parte anatomica rappresentata. Le serie di slices prodotte con apparecchiature differenti possono fornire diversi tipi di informazioni. Un uso sinergico di serie medicali differenti può quindi permettere di esaltare i pregi ed il contenuto informativo delle singole serie medicali, fornendo al medico il massimo supporto durante la diagnosi. In questo lavoro, vedremo come allineare tra loro, in lunghezza ed in risoluzione, diverse serie di slices allo scopo di renderle confrontabili e permetterne la fusione. L'immagine tridimensionale, risultato della fusione, sintetizza il contenuto informativo "saliente" delle singole immagini tridimensionali di partenza.

1. Introduzione

La straordinaria evoluzione delle tecnologie di acquisizione di immagini biomedicali consente ai medici che devono effettuare diagnosi cliniche di consultare diversi tipi di immagini.

Le immagini biomedicali acquisite da diverse modalità (come CT, PET, MRI, ecc.) presentano contenuti informativi spesso differenti, quali ad esempio morfologia (ad esempio, da un'immagine CT è possibile osservare se un tumore è penetrato o meno nella struttura ossea) o funzionale (ad esempio, da un'immagine PET è possibile definire la reale estensione di un tumore), e si differenziano per il livello di invasività,

per il costo, per le caratteristiche delle immagini acquisite, ad esempio in termini di risoluzione spaziale o temporale.

Ciascuna metodica di indagine presenta quindi peculiarità e limiti nel descrivere, dal punto di vista spaziale e temporale, lo stesso organo oggetto di studio. L'utilizzo di un unico tipo di immagine, spesso, non dà la possibilità al medico di determinare le possibili patologie che, invece, potrebbe individuare consultando contemporaneamente differenti tipi di immagini.

Un uso sinergico di immagini biomediche acquisite con modalità diverse permette, in tale ambito, di superare i limiti di ciascuna tecnica ed esaltarne i valori specifici consentendo al medico una più completa comprensione dei meccanismi fisiopatologici e produrre una più accurata definizione del quadro clinico. Per questo motivo, la fusione multimodale di immagini è un campo di ricerca estremamente attivo.

La fusione multimodale di immagini biomedicali permette di creare una singola immagine a partire da immagini ottenute con differenti modalità dello stesso soggetto e può consentire analisi e diagnosi addizionali in quanto estrae il contenuto informativo di due immagini differenti, sintetizzandolo nell'immagine fusa che produce. Tuttavia prima di effettuare la fusione è necessaria una co-registrazione delle immagini, dal momento che queste sono ottenute con apparecchiature o tempi diversi e quindi sono sicuramente non allineate.

In letteratura, è ben consolidata l'idea che l'analisi multi-risoluzione [1] (Multiresolution Analysis, o MRA) sia l'approccio che dà i risultati migliori nella fusione di immagini, questo per la sua capacità di decomporre un'immagine in più immagini a differenti risoluzioni che possono essere analizzate separatamente. La decomposizione Wavelet è il metodo che si adatta meglio all'approccio MRA applicato alle immagini, poiché essa riesce a separare un'immagine nelle sue componenti ad alta e bassa frequenza [2]. Ogni componente rappresenta le informazioni contenute ad una data scala. Il sistema di decomposizione è costruito in modo tale che è possibile ricostruire perfettamente il segnale originale, cioè, applicando la trasformata inversa alle componenti della decomposizione non c'è alcuna perdita di informazione nel sistema.

La seguente relazione descrive un algoritmo basato sulle Wavelet per la fusione multimodale di immagini mediche, in modo da estrarre le informazioni salienti da esse contenute.

2. Fusione multimodale di serie di slices biomedicali

In genere, non si ha a che fare con una singola immagine, in quanto esami clinici come la CT o la PET producono serie di immagini (slices). Ogni slice è la rappresentazione di una fetta del corpo del paziente con uno spessore compreso tra 1 e 10 mm. La risoluzione di una slice può variare tra i 64x64 e i 512x512 pixel. Il numero di slice varia a seconda dell'estensione del corpo da esaminare e dal particolare esame clinico che le ha prodotte.

Quindi, il processo di fusione di immagini andrà applicato ad ogni coppia di slices corrispondenti, così da generare una serie di slices fuse a partire da quelle di partenza.

Spesso, però, i differenti esami clinici disponibili producono serie di slices non omogenee tra loro, quindi in generale bisogna assumere che le serie di slices oggetto della fusione siano caratterizzate da proprietà diverse, sia in termini di numero di pixel costituenti le singole immagini, sia in termini di lunghezza della serie stessa (si possono avere ad esempio, 10 slices PET e 18 slices TAC).

Quindi, prima di procedere alla fusione delle serie di slices, bisogna far sì che tali slices siano normalizzate tra loro, per far sì che abbiano entrambe la stessa lunghezza (numero di slices) e che ogni slice abbia la stessa risoluzione (dimensione in pixel) .

In definitiva, il metodo di fusione proposto consiste in due passi:

1. Verifica delle condizioni iniziali: confronto fra la lunghezza e la risoluzione delle serie di slices da fondere e, solo nel caso in cui le due serie dovessero risultare differenti in uno dei due parametri suddetti, normalizzazione delle serie stesse.
2. Costruzione della serie fusa finale: applicazione della tecnica di fusione di immagini alle corrispondenti coppie di slices appartenenti alle due serie.

2.1 Verifica delle condizioni iniziali

Prima di procedere alla fusione delle serie di slices, bisogna verificare che esse abbiano la medesima lunghezza e che le slices di cui sono composte abbiano medesima risoluzione. Nel caso in cui almeno una delle condizioni suddette non sia verificata,

bisogna modificare opportunamente una delle due serie, in modo tale che alla fine esse abbiano la stessa lunghezza e risoluzione.

Se la risoluzione delle slices di una delle due serie è più piccola rispetto alla risoluzione delle immagini dell'altra serie, bisogna aumentarne la risoluzione in modo che le slices delle due serie abbiano entrambe la stessa risoluzione. Ovviamente, la normalizzazione delle risoluzioni potrebbe anche essere fatta riducendo la risoluzione della serie a risoluzione maggiore, però la riduzione di risoluzione di un'immagine comporta una perdita di informazioni rispetto all'immagine originale.

L'aumento di risoluzione di un'immagine comporta l'aumento del numero dei pixel costituenti l'immagine stessa, quindi, anche se non si ha perdita di informazioni nel processo, bisogna comunque determinare il valore dei nuovi pixel che formeranno l'immagine finale.

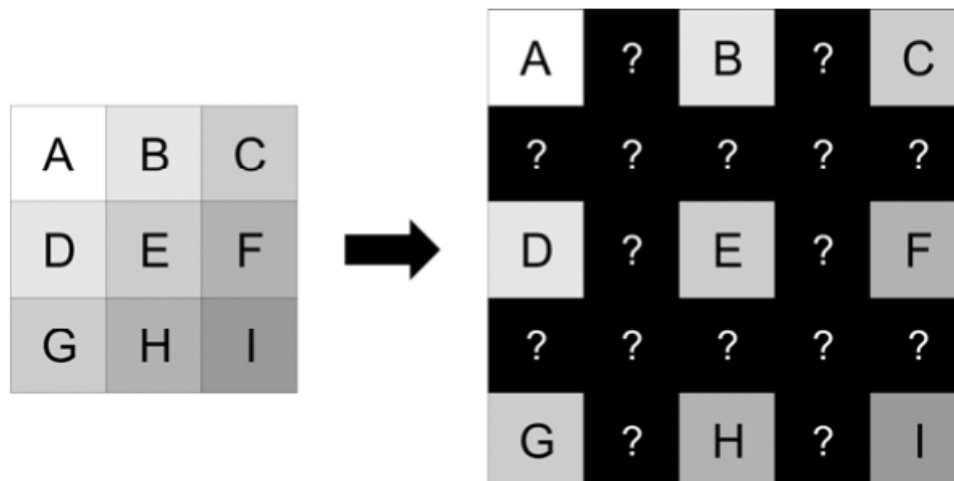


Figura 2.1 – Passaggio da bassa ad alta risoluzione

Nell'immagine ad alta risoluzione, la stima del valore dei pixel mancanti viene fatta utilizzando delle tecniche di interpolazione che sfruttano il valore dei pixel noti. Ovviamente, il risultato è comunque frutto di un'approssimazione quindi, è lecito aspettarsi degli artefatti nell'immagine ad alta risoluzione [3].

Se le due serie di slices da fondere hanno lunghezza diversa, risulta necessario aumentare il numero di slices della serie con meno elementi in modo che le due serie abbiano la stessa lunghezza al termine del procedimento. Per fare ciò, si ricostruiscono le slices mancanti della serie con meno immagini interpolando le slices di quest'ultima. Quindi, in definitiva, si potrebbe avere la necessità di ricostruire non solo le parti mancanti di una slice (i pixel ignoti generati dal suo aumento di risoluzione), ma anche intere slices della stessa serie (in modo da equipararne la lunghezza all'altra serie).

La tecnica che abbiamo utilizzato si basa sul concetto che una serie di slices mediche 2D può essere vista come un'unica immagine tridimensionale, dove il pixel di coordinate (x,y,z) corrisponde al pixel di coordinate (x,y) della slice individuata da z . Quindi, a partire da un'immagine 3D di dimensioni (m,n,k) , si vuole ricostruire, tramite un'interpolazione tridimensionale, l'immagine 3D di dimensioni (a,b,c) , dove: (a,b) è la risoluzione che deve avere ogni slice della serie al termine dell'elaborazione e c è la lunghezza della serie che si vuole ottenere. Quindi, si effettua un aumento della risoluzione tridimensionale dell'immagine 3D iniziale (la serie di slices) portandola alla risoluzione spaziale opportuna (quella dell'altra serie di slices).

La scelta della particolare tecnica di interpolazione dipende dal rapporto tra precisione e velocità di calcolo che si vuole ottenere.

I metodi di interpolazione possono essere raggruppati in due principali categorie: *adattativi* e *non adattativi*. I metodi adattativi utilizzano tecniche diverse a seconda del tipo di regione che stanno interpolando, mentre i metodi non adattativi trattano tutti i pixel allo stesso modo. Diversi metodi non adattativi possono essere combinati in un unico metodo adattativo. I metodi di interpolazione possono essere più generalmente classificati come *lineari* e *non lineari*. I metodi lineari utilizzano filtri lineari spazialmente invarianti che, applicati all'immagine a bassa risoluzione, generano i pixel ad alta risoluzione. I metodi non lineari utilizzano un processo di ottimizzazione non lineare vincolato da alcune caratteristiche (*features*) dell'immagine a bassa risoluzione. Data la loro semplicità ed efficienza, i metodi lineari sono quelli maggiormente utilizzati. Fanno parte di questa categoria i metodi *nearest neighbor*, *lineare* e *cubico*.

Nearest neighbor: noto anche come *pixel replication*, è in assoluto il metodo di interpolazione più semplice e rapido. Questo metodo si limita a porre il valore di ogni pixel sconosciuto uguale al valore del pixel noto più vicino. L'immagine interpolata appare come se ogni pixel fosse cresciuto proporzionalmente al fattore di ingrandimento.

Lineare: nella versione 2D (bilinare), stima il valore del pixel sconosciuto facendo la media dei valori dei quattro pixel noti diagonalmente adiacenti. Da un punto di vista geometrico, corrisponde ad utilizzare un piano interpolante che passa per i quattro pixel noti. Il risultato visivo di questo tipo di interpolazione è un'immagine

tanto più sfocata quanto maggiore è il fattore di ingrandimento. Nella versione 3D (trilineare), viene fatta la media degli otto pixel noti diagonalmente adiacenti.

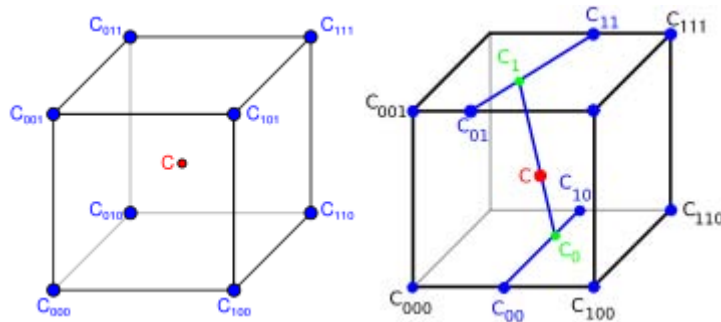
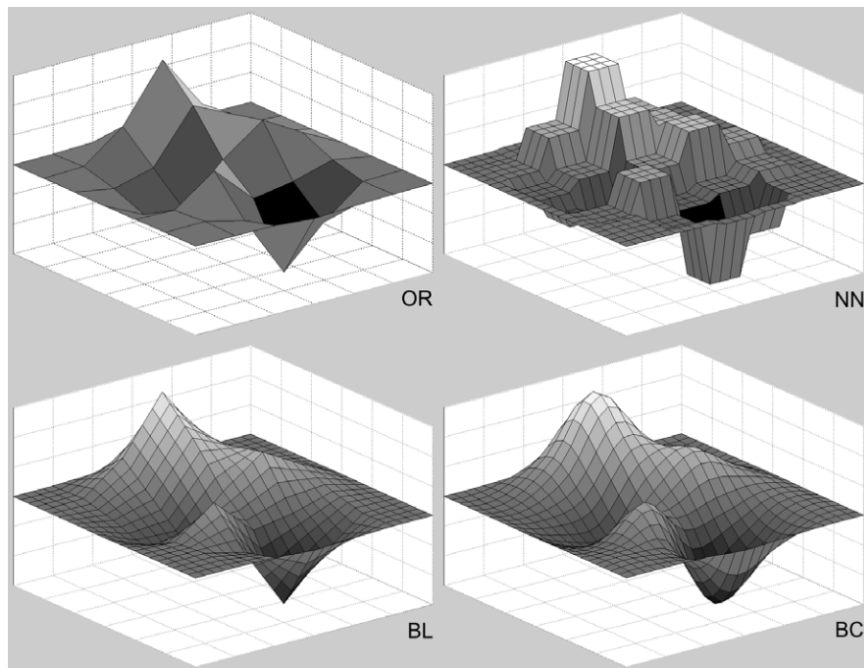


Figura 2.2 – Interpolazione trilineare

Cubico: utilizza lo stesso principio del metodo lineare ma, al posto della media dei pixel adiacenti noti, si utilizza un polinomio interpolante bicubico (versione 2D) vincolato ai valori dei 16 pixel noti che circondano il pixel da interpolare in una griglia 4×4 . Questo metodo genera immagini più nitide rispetto al metodo bilineare. Nella versione 3D si utilizza un polinomio interpolante tricubico, vincolato ai valori dei 64 pixel noti che circondano il pixel ignoto in uno spazio $4 \times 4 \times 4$. L'ottimo rapporto tra qualità e complessità computazionale ne fanno uno dei metodi di interpolazione più utilizzati.



Legenda

(in altezza vengono riportati i valori di intensità dei pixel)

- *OR* = immagine sottocampionata di partenza;
- *NN* = nearest neighbor (pixel replication);
- *BL* = bilineare;
- *BC* = bicubico;

Figura 2.3 – Comparazione tridimensionale di alcuni metodi lineari (ingrandimento 4x)

Ci sono molte altre tecniche di interpolazione che prendono in considerazione molti più pixel durante il procedimento, ma comportano una complessità computazionale molto maggiore rispetto a quelle viste precedentemente. Tali tecniche includono funzioni spline e sinc, riuscendo a conservare la maggior parte del contenuto informativo dell'immagine dopo un'interpolazione. Esse risultano particolarmente efficaci quando l'immagine iniziale richiede elaborazioni multiple tipo rotazioni e distorsioni effettuate in passi successivi.

interpolation	kernel size	MRE (%)	error ratio	run-time ratio
nearest neighbor	1×1	3.397	1.712	1.00
linear	2×2	2.280	1.149	2.81
B-spline	4×4	2.637	1.330	10.58
cubic	4×4	2.005	1.011	10.24
cubic	6×6	2.010	1.013	20.03
truncated Sinc	6×6	12.063	6.080	20.08
windowed sinc	6×6	1.984	1.000	20.19

Tabella 2.1 – Confronto tra alcuni metodi di interpolazione

Nel nostro caso, invece, abbiamo bisogno semplicemente di un'elaborazione in un singolo passo, cioè un aumento della risoluzione tridimensionale dell'immagine iniziale. Quindi, possiamo utilizzare un metodo relativamente poco oneroso che, per il nostro caso, risulta comunque adeguatamente accurato. In particolare, la scelta è caduta sull'interpolazione di tipo trilineare.

2.2 Fusione delle coppie di slices

Una volta verificate le condizioni iniziali, ed eventualmente eseguite le operazioni di interpolazione, si può dunque procedere alla effettiva fusione delle due serie, effettuando la fusione tra ogni slice della prima serie con la corrispondente slice della seconda serie.

L'algoritmo consiste in una tecnica a finestra che combina tra loro le componenti a bassa frequenza (approssimazioni) di un'immagine e le componenti ad alta frequenza (dettagli) di un'altra immagine utilizzando rispettivamente la massima varianza ed il massimo modulo come indici di qualità. Sono considerate informazioni significative il cambiamento di contrasto dei pixel nel caso delle approssimazioni ed i contorni nel caso dei dettagli.

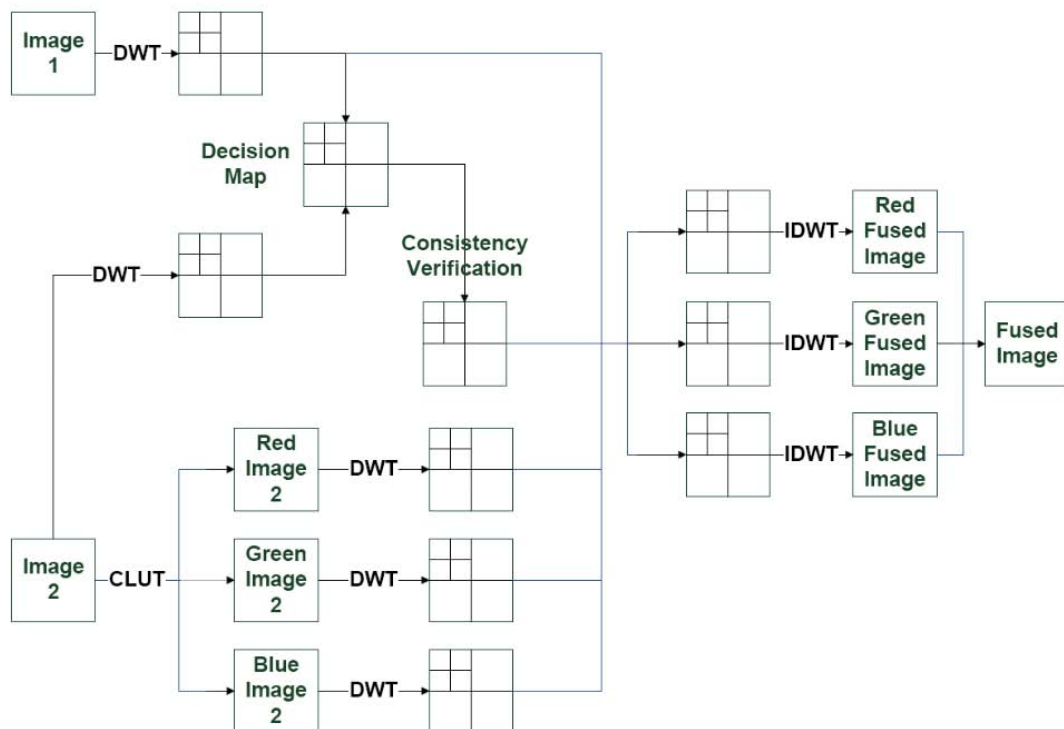


Figura 2.4 – Schema dell’algoritmo di fusione proposto

Sinora, molte altre tecniche di fusione basate su Wavelet sono state proposte [4, 5, 6]. La soluzione presentata in questa relazione differisce da esse per due ragioni basilari:

- i. è un approccio completamente selettivo (per esempio, i coefficienti di approssimazione non vengono combinati attraverso la loro media).
- ii. una delle immagini è colorata prima che il procedimento della fusione cominci; quindi, è stato affrontato il problema di combinare separatamente i canali RGB, con lo scopo di dare al medico la possibilità di distinguere, nell’immagine fusa ottenuta, se un pixel provenga da un’immagine o dall’altra.

Partendo dal presupposto che tutte le immagini da fondere siano in scala di grigio, la tecnica di fusione che deve essere applicata ad ogni coppia di slices può essere suddivisa in cinque passi fondamentali. Siano *Image1* ed *Image2* le due slice da fondere:

1. *Image2* viene colorata tramite una Color Look-up Table (CLUT), una tecnica che trasforma un pixel in scala di grigio in un pixel RGB consultando una tabella nota. Dopodichè, dopo aver fissato un livello di decomposizione, la trasformata Wavelet discreta 2-D decompone *Image1* e *Image2* nelle sue componenti RGB. Si ottengono così cinque matrici di coefficienti diverse.
2. Vengono calcolate e comparate le varianze dei coefficienti contenuti nelle finestre che scorrono le decomposizioni di *Image1* e di *Image2*. Detta S l'immagine sorgente, W la dimensione della finestra, K il livello di decomposizione ed a la matrice dei coefficienti di approssimazione, la varianza può essere così definita:

$$var_S^{(K)}(W) = \frac{1}{|W|} \sum_{(i,j) \in W} \left(a_S^{(K)}(i,j) - \frac{1}{|W|} \sum_{(i,j) \in W} a_S^{(K)}(i,j) \right)^2$$

I coefficienti di approssimazione dell'immagine fusa sono tutti i pixel centrali delle finestre con il massimo valore di varianza. In particolare, viene creata un'immagine binaria (detta *decision map*) in cui i valori 0 o 1 indicano se il coefficiente finale sarà estratto dalla prima o dalla seconda immagine.

3. Per quanto riguarda i coefficienti di dettaglio, vengono calcolate le somme dei moduli dei coefficienti contenuti in una finestra mobile di dimensione W , nel modo seguente:

$$mod_S^{(K)}(W) = \sum_{(i,j) \in W} |d^{(K)}(i,j)|$$

dove d è la matrice dei coefficienti di dettaglio. La decision map viene completata considerando tutti i pixel centrali delle finestre aventi modulo massimo.

4. Viene eseguita una verifica di consistenza della decision map attraverso una finestra mobile di dimensione W che scorre la decision map per rilevare la presenza di rumore, che è caratterizzato dalla scelta di estrarre un pixel da un'immagine, mentre i pixel più vicini (cioè i pixel della finestra tranne il pixel

centrale) sono estratti dall'altra immagine. Soltanto in questo caso, l'algoritmo invertirà il valore di tale pixel della decision map.

5. I canali RGB dell'immagine fusa decomposta vengono determinati estraendo i coefficienti dalle decomposizioni Wavelet dell'immagine 1 e dai canali RGB dell'immagine 2, secondo le informazioni della decision map. Dopodichè, ai canali RGB così ottenuti viene applicata la trasformata Wavelet discreta 2-D inversa. Si otterrà così l'immagine RGB fusa dalle due di partenza.

3. Risultati sperimentali

Questo paragrafo riporta un confronto tra la tecnica di fusione proposta ed un'altra tecnica che effettua la media dei coefficienti di approssimazione. I risultati sono stati valutati usando due differenti indici di qualità: la mutua informazione [7], spesso utilizzata anche per stimare la registrazione di immagini mediche, e l'indice proposto da Wang e Bovik [8]. L'algoritmo è stato implementato in C++ e sono state utilizzate due librerie open source: ITK [9], per l'elaborazione di immagini mediche nel formato standard DICOM [10], e VTK [11], per la visualizzazione delle immagini stesse. I dati considerati per il test sono due serie di immagini (CT e PET) di 10 slices ciascuna. La risoluzione delle immagini CT è di 512x512 pixel, mentre le immagini PET hanno risoluzione 128x128 pixel. Quindi, avendo le due serie medesima lunghezza, per applicare la nostra tecnica bisogna soltanto aumentare la risoluzione delle immagini PET e portarla a 512x512 pixel.

Slice	<i>Mutual Information</i>		<i>Wang and Bovik</i>	
	Mean	Variance	Mean	Variance
1	0.383356	0.395590	0.383717	0.403864
2	0.385227	0.401047	0.384714	0.406014
3	0.392685	0.413141	0.388363	0.407864
4	0.391598	0.413616	0.387095	0.406453
5	0.392148	0.412260	0.387202	0.403259
6	0.394681	0.414405	0.389981	0.404458
7	0.399621	0.415722	0.387454	0.400452
8	0.385982	0.405891	0.384140	0.396303
9	0.402599	0.419474	0.389815	0.401747
10	0.400288	0.418346	0.382649	0.396750

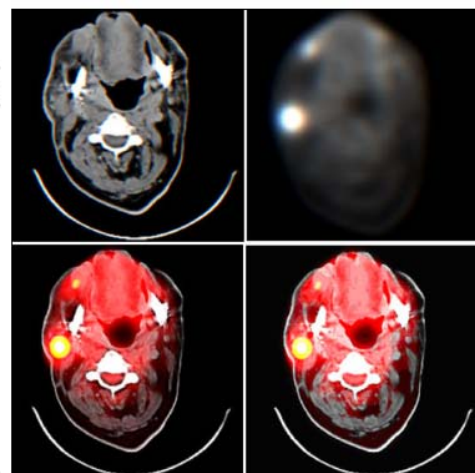


Figura 2.5 – Risultati numerici ed, in senso orario, una CT, una PET, fusioni con varianza e media

Le immagini sono state decomposte tramite Wavelets Daubechies di livello 20 con due livelli di decomposizione, mentre la dimensione della finestra mobile scelta è 5x5.

La figura 2.5 mostra una valutazione qualitativa e quantitativa. Come si vede, sia la mutua informazione che l'indice di Wang e Bovik danno risultati incoraggianti. Inoltre, l'immagine fusa tramite l'approccio proposto risulta più accurata e le caratteristiche delle immagini sorgenti sono più distinguibili di quella che si ottiene con il criterio classico basato sulla media delle approssimazioni.

Bibliografia

1. Piella, G., "A General Framework for Multiresolution Image Fusion: From Pixels to Regions". *Information Fusion* 4, 259–280 (2003)
2. Chipman, L.J., Orr, T.M., Lewis, L.N., "Wavelets and Image Fusion". *IEEE Transactions on Image Processing* 3, 248–251 (1995)
3. Thévenaz, P., Blu, T., Unser, M., "Image Interpolation and Resampling" . *Handbook of medical imaging*, ISBN: 0-12-077790-8, 2000.
4. Berbar, M.A., Gaber, S.F., Ismail, N.A., "Image Fusion Using Multi- Decomposition Levels of Discrete Wavelet Transform". *Visual Information Engineering*, 294–297 (2003)
5. Pajares, G., de la Cruz, J.M., "A Wavelet-based Image Fusion Tutorial". *PR(37)* 9, 1855–1872 (2004)
6. Qu, G., Zhang, D., Yan, P., "Medical Image Fusion by Wavelet Transform Modulus Maxima". *Optics Express* 9(4), 184 (2001)
7. Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A., "Mutual Information Based Registration of Medical Images: A Survey". *IEEE Transactions on Medical Imaging* 22(8), 986–1004 (2003)
8. Wang, Z., Bovik, A.C., "A Universal Image Quality Index". *IEEE Signal Processing Letters* (2002)
9. <http://www.itk.org/>
10. <http://medical.nema.org/>
11. <http://public.kitware.com/VTK/>