

Tecniche di rappresentazione di immagini orientate alle strutture

Giovanni Iacovitti

INFOCOM Dpt., University of Rome "La Sapienza"

email: [gjacob@infocom.ing.uniroma1.it](mailto:gjacov@infocom.ing.uniroma1.it)

Sommario

La presente relazione illustra sinteticamente il contributo alla attività di ricerca dell'unità di Roma I. Dopo una introduzione al tema specifico e alle metodologie seguite, è dapprima presentato in sintesi il quadro strumentale matematico entro cui si è svolto il lavoro (wavelet armoniche circolari), che è stato anche alla base di ricerche precedenti o che sono confluite nella ricerca in oggetto. Sono quindi trattati in qualche dettaglio alcuni dei problemi che hanno impegnato l'unità (controllo della qualità delle immagini, astrazione della rappresentazione mediante bordi). Segue una rapida rassegna di altri problemi affrontati, a partire dal riconoscimento e localizzazione automatica di strutture e dall'analisi e sintesi di tessiture, fino alla segmentazione. La relazione è corredata da alcune immagini che esemplificano i risultati ottenuti.

1. Introduzione

La citazione dei meccanismi di visione naturale non è infrequente nel mondo della elaborazione delle immagini. Se tuttavia il ricorso ad analogie con modelli di percezione è stato utile in problemi che riguardano la qualità di presentazione a fruitori umani (ad esempio nella costruzione di procedimenti di compressione), non altrettanto scontato appare l'uso questa metodologia quando occorre affrontare problemi di analisi o di interpretazione automatica del contenuto delle immagini. Infatti, la conoscenza sui sistemi di visione dei mammiferi, ed in particolare dell'uomo, è tuttora assai limitata per quanto riguarda i livelli più alti, dove hanno luogo i processi cognitivi e interpretativi. È stato anzi autorevolmente argomentato come sia addirittura improprio, in linea di principio, stabilire modelli algoritmici per tali fenomeni.

Ciò nonostante, in considerazione della incomparabile efficienza dei sistemi naturali di visione, appare particolarmente interessante, anche per quanto riguarda i sistemi di visione artificiale, trarre ispirazione da quanto di noto esiste su quelli, per lo meno per i livelli più elementari (la formazione delle immagini e la loro organizzazione in flussi informativi verso i centri nervosi centrali).

Tra le conoscenze più consolidate attorno alla visione dei mammiferi vanno annoverate certamente *la rappresentazione a risoluzione differenziata e la esistenza di operatori di estrazione di bordi con la loro orientazione.*

I recenti sviluppi matematici riguardanti la rappresentazione dei segnali a multirisoluzione, ed in particolare l'analisi delle immagini nel dominio trasformato di particolari wavelet che possono essere considerate come operatori di estrazione dei bordi [1], costituiscono quindi elementi di ovvio interesse anche per l'analisi automatica delle immagini, oltretutto, naturalmente, per il controllo di qualità nei procedimenti di presentazione.

Ciò illustra i motivi che hanno suggerito l'orientamento impresso al contributo di ricerca da parte dell'unità di Roma I.

Nel campo dei sistemi di comunicazione multimediale, cresce infatti l'esigenza di disporre di ausili per la consultazione e la composizione di documenti di immagine. A differenza dei testi alfanumerici, e analogamente a quanto avviene nei documenti audio, nelle immagini non è infatti semplice stabilire una connessione tra l'espressione sensibile del documento ed elementi del suo contenuto informativo in termini di significato, ed occorrono ausili basati su metodi di analisi automatica che permettano, più o meno rudemente, di selezionare classi di immagini in base a caratteristiche in qualche modo correlate al loro contenuto semantico. Al tempo stesso, la vasta articolazione in termini di larghezza di banda disponibile all'utente, pone l'esigenza di disporre di mezzi assai flessibili e immuni da artefatti per l'erogazione dell'informazione visiva a diversi livelli di *bitrate*.

In tale contesto, l'obiettivo della ricerca è stato quello di definire e verificare tecniche di rappresentazione che consentissero ad un tempo di soddisfare da una parte ai requisiti di controllo di qualità visuale, e, dall'altra, ad esigenze di descrittibilità strutturale dell'immagine stessa. Per le ragioni ora accennate, la ricerca si è articolata attraverso un approfondimento sulle rappresentazioni multirisoluzione basate su wavelet orientate ai bordi, e susseguentemente sulla definizione di schemi per il controllo di qualità e di estrazione di caratteristiche di contenuto a partire da tali rappresentazioni.

L'obiettivo comprendeva anche tecniche di analisi delle tessiture che, rispetto ai bordi, costituiscono nelle immagini l'elemento complementare, ad essi non facilmente riconducibile per quanto si conosce sui meccanismi di percezione naturale.

2. Le wavelet armoniche circolari

Le tecniche di rappresentazione delle immagini attraverso i bordi, non sono nuove. Esse sono state prese in considerazione più volte in modo più o meno sistematico nel contesto degli studi per la codifica di immagini ad alto tasso di compressione.

Sul piano formale, un contributo determinante è stato fornito nel contesto delle tecniche di espansione di immagine mediante wavelet, attraverso l'uso congiunto di due wavelet monodimensionali, originate rispettivamente da operatori differenziali orizzontali e verticali. Questi operatori, ottenuti come derivate di gaussiane, costituiscono estrattori di bordi verticali e rispettivamente orizzontali [1].

Una formulazione simile ma più compatta è quella basata su una sola wavelet bidimensionale complessa, derivata da una mother wavelet costituita da una funzione armonica circolare (CHF) di ordine 1. In breve, una CHF di ordine n è una funzione polarmente separabile, con andamento angolare armonico di periodo $n \cdot 2\pi$. Per $n=1$ si

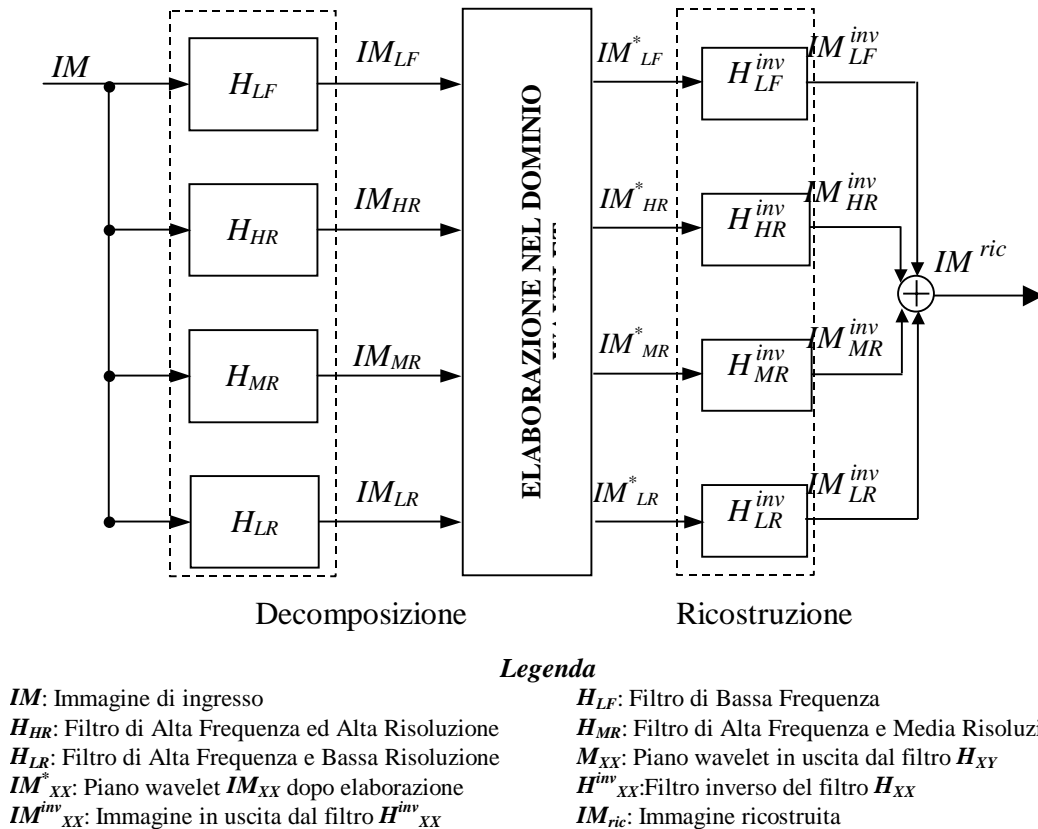


Figura 1. Diagramma a blocchi Analisi-Sintesi

ottiene un operatore che estrae dall'immagine i cosiddetti *bordi complessi*, caratterizzati dalla loro intensità (modulo) e direzione (fase). La corrispondente wavelet armonica circolare (CHW) di ordine 1 è adatta dunque a rappresentare una immagine attraverso coefficienti wavelet che indicano di fatto la esistenza di bordi complessi a diversi livelli di risoluzione. Tale schema si è rivelato estremamente flessibile per gli scopi di controllo di qualità di restituzione da una parte e di rappresentazione strutturale dall'altra, che sono gli obiettivi fondamentali del programma dell'unità di ricerca. Lo schema di base utilizzato nelle diverse fasi applicative del programma quello descritto nella figura 1.

Lo stadio di analisi (o di decomposizione) consiste in un banco di filtri: esso comprende una via che estrae la componente a bassa frequenza e tre vie di estrazione di bordi a tre livelli di risoluzione (bassa, media, alta). Le uscite costituiscono i piani di rappresentazione wavelet nel dominio spazio-scala. In questo dominio vengono effettuate le operazioni di controllo di qualità e di analisi di contenuto di tipo *multiscala*, mediante schemi di elaborazione diversi, come descritto in seguito.

Lo stadio di sintesi, utilizzato nelle applicazioni di controllo di qualità di restituzione, consiste in una combinazione di filtri che, in assenza di elaborazione nel dominio wavelet, ricostituiscono l'immagine originale. In altre parole, l'immagine può essere costruita dai suoi bordi complessi, assieme ad una componente di bassa frequenza.

Uno schema di analisi/sintesi pure fondato sull'uso di funzioni armoniche circolari è quello che fa uso non di diversi livelli di risoluzione, ma di diversi *ordini di armonicità* n . In tal caso, l'immagine viene decomposta in diverse immagini complesse ognuna delle quali indica l'esistenza di contributi di configurazioni diverse, a partire dai bordi ($n=1$), linee ($n=2$), incroci ortogonali ($n=4$) e così via. Questa decomposizione, dà luogo, nella sua forma particolare derivata da funzioni gaussiane differenziate, alla cosiddetta trasformata di *Laguerre-Gauss*. Tale trasformata è interessante per le sue cosiddette caratteristiche di *self-steerability*: ciò significa che nel dominio trasformato la rotazione di una struttura attorno ad un determinato centro di rotazione corrisponde alla moltiplicazione per coefficienti complessi della trasformata calcolata nel centro di rotazione. Questa caratteristica rende la trasformata di Laguerre Gauss particolarmente utile per l'analisi *multiorientazione* delle immagini.

Infine, l'utilizzazione congiunta di differenti ordini di armonicità e di diverse scale conduce ad uno strumento di analisi assai completo, costituito dalla cosiddetta piramide armonica circolare (CHP) che consente ad un tempo l'analisi *multiscala e multiorientazione* delle immagini.

Lo studio matematico di tali schemi ha consentito di determinarne alcune utili proprietà e di definire i parametri e le modalità più convenienti di utilizzo nelle applicazioni. Di seguito, vengono descritte le applicazioni più interessanti che sono state sviluppate o affinate nel corso della ricerca.

3. Controllo di qualità [3]

Negli ultimi anni sono stati proposti diversi schemi di codifica di immagine a fedeltà adattativa nello spazio-tempo, per utilizzare nel modo più conveniente la larghezza di banda disponibile per il trasferimento dell'immagine in relazione a criteri di importanza visuale. Lo scopo è quello di *amministrare* efficientemente il livello di qualità di restituzione in relazione alla attitudine percettiva dell'utente. I limiti percettivi ai livelli più bassi (struttura dei campi percettivi, effetti di mascheramento, etc.) sono già ampiamente sfruttati nelle tecniche di compressione video per ridurre la ridondanza di rappresentazione del segnale.

Il sistema visivo presenta peraltro caratteristiche di limitazione percettiva nello spazio tempo assai più complesse e non ancora ben identificate e comprese.

Per costruire una percezione soggettiva di grande nettezza, estensione angolare ed immediatezza, il sistema visivo dei mammiferi sembra utilizzare meccanismi raffinati che comprendono in primo luogo una distribuzione differenziata delle risorse di banda spaziale e temporale nella mappa della retina. Al centro della retina (fovea) si dispone della banda spaziale più ampia, che consente di percepire particolari con elevata acuità visuale ($0.5'$). Al di fuori di questa zona, che si estende per appena 5° , la banda spaziale decresce rapidamente. Tale diminuzione è però accompagnata da una crescente banda temporale, che permette di cogliere con rapidità ed evidenza variazioni di luce nella zona periferica del campo visivo. Tali caratteristiche potrebbero essere utilizzate per modulare secondo opportune strategie la

riproduzione delle immagini, concentrando le risorse di banda spaziale e temporale dove più richiesto. E' ovvio però che occorrerebbe conoscere la direzione istantanea di puntamento dell'asse foveale. Il problema è complicato dal fatto che alla formazione della percezione sembrano contribuire decisamente aspetti legati alla strategia volontaria ed involontaria (saccadi) di esplorazione della scena, e altri probabili meccanismi non lineari che producono apparente definizione anche in zone eccentriche rispetto all'asse foveale.

Tuttavia queste difficoltà non vanificano del tutto la possibilità di sfruttare le differenze di risoluzione spazio tempo del sistema visivo. Per risolvere almeno sul piano statistico tale problema sono state proposte tecniche di determinazione strumentale delle cosiddette "regioni di interesse", sulla base di diversi criteri. Vi sono molte applicazioni (come la telesorveglianza) dove la zona di interesse è indicata esplicitamente dall'utente, oppure dettate dall'applicazione stessa (localizzazione mediante strumentazione esterna). Nelle applicazioni più usuali di *immagini* di TV commerciale e fiction occorre poi considerare come le stesse tecniche di ripresa siano concepite generalmente in modo da mantenere il centro di interesse dell'utente a centro schermo. Queste circostanze lasciano pensare alla possibilità effettiva di migliorare in certa misura l'uso della banda disponibile, assecondando l'attitudine naturale a recepire maggiori dettagli sull'asse foveale.

Progressive esperienze in tal senso sono state condotte già da tempo dal gruppo di ricerca. Esse hanno portato ad un procedimento consistente in uno schema di **pre-processing** in trasmissione ove si applica lo schema di analisi sintesi a multirisoluzione sopra descritto per modulare la risoluzione delle immagini trasmesse secondo lo schema di visione foveale (posizionato di norma a centro schermo). Successivamente, il processo di modulazione della risoluzione è stato via via sofisticato tenendo conto della evoluzione dinamica delle immagini. In particolare è stato amplificato l'effetto di decadimento periferico della risoluzione in corrispondenza di rapide variazioni della scena. Questo fatto ha permesso di attutire sensibilmente i picchi informativi della sorgente di immagine, permettendo così di agevolare l'amministrazione del flusso informativo erogato senza la generazione di artefatti disastrosi.

Nella presente ricerca, tale procedimento è stato completato con la definizione e la sperimentazione iniziale di uno schema di **post-processing** dal lato ricezione.

Tale completamento è stato dettato dalla necessità di ridurre la percezione di diminuzione della risoluzione in periferia, comunque visibile (nei formati di presentazione più ridotti) come una sfocatura dell'immagine.

Lo schema di post-processing è finalizzato al recupero apparente di risoluzione nelle zone affette da perdita di risoluzione (in questo caso intenzionale). Le tecniche convenzionali usate per questo scopo consistono in una esaltazione delle alte frequenze spaziali mediante operatori lineari di derivazione. Tali tecniche non sono applicabili nel nostro caso anche a causa della presenza degli errori di codifica che si presentano come disturbi strutturati (blocking, etc.) che rischiano di essere fortemente evidenziati. La tecnica di recupero impiegata è invece non lineare e viene condotta nel dominio wavelet CHW (al centro dello schema di analisi/sintesi di fig.1). In accordo a concetti ampiamente discussi in [1] il recupero di risoluzione consiste in linea di principio nella *equalizzazione lungo la direzione della scala* del modulo dei coefficienti wavelet. A seguito di prove soggettive condotte su

immagini degradate in risoluzione per differenti cause, sono stati identificati i parametri di una non-linearità senza memoria di tipo polinomiale di 2° grado ($y(x) = \alpha \cdot x^2 + \beta \cdot x + \delta$) che fornisce risultati di recupero superiori e quelli forniti dalle tecniche convenzionali lineari, specialmente in presenza di disturbi. Un esempio di recupero di nitidezza apparente è mostrato nella figura 2). Vale la pena di sottolineare che si tratta di un recupero solo apparente di dettagli, ottenuta con tecniche non lineari. A differenza dei metodi fondati sulla equalizzazione spettrale (rifocalizzazione, edge enhancement), questo metodo è applicabile nei confronti di una ampia classe di cause di perdita di risoluzione.



Figura 2a. Immagine sfocata



Figura 2b. Immagine restaurata

Il metodo così definito è stato quindi applicato a *sequenze di immagini* di tipo videotelefonico, sottoposte a preprocessing di tipo foveale e quindi codificate in H263. E' stato necessario in questo caso applicare nel dominio wavelet anche un procedimento di de-blocking, consistente nella soppressione di bordi in corrispondenza alle discontinuità interblocco, e conseguente interpolazione spaziale dei coefficienti wavelet soppressi. I risultati del procedimento complessivo sono visibili nelle figure 3a) e 3b) che riportano fotogrammi comparati di sequenze videotelefoniche a parità di *bitrate* (circa 30 Kbit/sec).

4. Astrazione della rappresentazione tramite bordi [11]

L'analisi delle immagini sulla base dei contenuti richiede l'esame di caratteristiche dell'immagine stessa che siano sufficientemente indicative dei contenuti. Alcuni esempi sono il colore dominante, l'esistenza di strutture fortemente direzionali, la presenza di particolari pattern, etc. Per disporre di mezzi più potenti di discriminazione occorre arrivare a livelli di analisi di tipo sintattico. Questi richiedono però una rappresentazione dell'immagine di livello sufficientemente astratto da poter essere tradotta significativamente in costrutti simbolici interpretabili.

Un passo fondamentale in questo senso è costituito dalla riduzione della descrizione dell'immagine in termini di primitive grafiche semplici, tipicamente segmenti di retta in luogo dei bordi. Tuttavia, sebbene la descrizione dell'immagine in termini di segmenti

associati ai bordi presenti nell'immagine sia da considerare classica, essa si realizza in pratica non senza difficoltà.

Nella presente ricerca è stato impiegato lo schema di analisi di fig. 1 come stadio preliminare di un procedimento di rappresentazione semplificata dell'immagine. Il procedimento prevede dapprima una soppressione di bordi scarsamente significativi per ampiezza o estensione, ed in una approssimazione dei rimanenti mediante spezzate costituite da segmenti rettilinei. Ciò porta ad una descrizione parametrica dell'immagine in termini di segmenti che rappresentano allineamenti di coefficienti wavelet che corrono lungo i contorni degli oggetti. L'immagine può essere quindi ricostruita a partire da segmenti approssimanti i bordi, con un livello di qualità dipendente dal numero dei segmenti utilizzati.



Figura 3a. H263 (sinistra), H263+ PrePost (destra)



Figura 3b. H263 (sinistra), H263+ PrePost (destra)

Questa rappresentazione parametrica corrisponde, attraverso lo schema di sintesi, ad una immagine semplificata nel dominio spaziale, come esemplificato in fig. 4. Nella sostanza dunque, il procedimento di astrazione coincide con un metodo di codifica basato su segmenti.

Le maggiori difficoltà tecniche di tale metodologia di rappresentazione parametrica si trovano quando esse sono applicate a sequenze di immagini. Procedendo nel lavoro di ricerca, per tale scopo è stato necessario stabilire una associazione tra segmenti corrispondenti in immagini consecutive mediante una tecnica inusuale di motocompensazione. Come risultato primario, ciò ha portato alla determinazione di traiettorie di segmenti rettilinei piuttosto profonde nello spazio tempo (oltre l'orizzonte della decina di quadri), e quindi all'individuazione di primitive geometriche persistenti. Si pensa



Figura 4a. Originale



Figura 4b. Versione parametrica semplificata dei bordi complessi



Figura 4c. Immagine costruita mediante i bordi complessi semplificati

che tali elementi descrittivi invarianti lungo la sequenza potranno essere utilmente impiegati a livelli superiori di analisi sintattica per la classificazione di sequenze di immagini in base al contenuto.

Nell'ambito stesso della ricerca, sono stati condotti alcuni esperimenti. Sono state provate in particolare alcune procedure di segmentazione temporale delle sequenze video in base alla persistenza di strutture nella scena, procedure che presentano interessanti caratteristiche di resistenza rispetto a movimenti della camera di ripresa e a operazioni di zooming.

5. Ulteriori attività

Di seguito sono brevemente riportati ulteriori risultati conseguiti nell'ambito della ricerca, con esempi illustrativi.

Riconoscimento e localizzazione di strutture [6,7,9,10]

In collaborazione con l'unità di ricerca di Roma III sono state ulteriormente sviluppate le tecniche di classificazione mediante detezione di preassegnate strutture (*pattern di riferimento*) nella scena esaminata, basata sullo sviluppo di Laguerre-Gauss. Utilizzando queste funzioni per rappresentare sia il pattern di riferimento che le immagini in esame, sono stati definiti procedimenti di classificazione e riconoscimento in base a criteri sia statistici che deterministici, e sono state formulate soluzioni ottimali. Riguardo ai criteri statistici, è stato considerato il problema di rivelazione di distribuzioni spaziali di caratteristiche (features) di immagine. Riguardo ai criteri deterministici, è stato considerato lo sviluppo di oggetti campione attorno ad un punto convenzionale di riferimento. La comparazione delle componenti di sviluppo rispetto ad ogni punto dell'immagine esaminata permette di rivelare la presenza di copie dell'oggetto campione, indipendentemente dalla rotazione e dalla scala, nonché di stimare posizione e orientamento. In particolare, è stata definita una strategia basata sull'uso di rapporti di verosimiglianza generalizzati susseguenti a stime della rotazione del campione per ogni ipotesi di posizionamento sul piano dell'immagine esaminata. Tale applicazione è descritta in dettaglio nella relazione [9] riportata nella presente raccolta. Tra gli sviluppi si annovera una procedura di moto-compensazione che tiene conto anche di componenti di rotazione nel movimento tra quadri di una sequenza video.

Classificazione e sintesi di tessiture [5,8,12,13,]

In collaborazione con l'unità di ricerca di Roma III una tecnica di analisi e sintesi di tessiture basate sulla decomposizione di tipo Bussgang tecnica, già sviluppata in attività di ricerche precedentemente avviate, è stata largamente analizzata ed estesa al caso delle tessiture a colori. È stato utilizzato in sostanza un modello statistico di tessitura basato su processi Gaussiani sottoposti a trasformazioni non-lineari con memoria (costituiti da una coppia di filtri lineari connessi tramite non-linearità senza memoria. Un esempio di risultati è riportato in fig.5.



Figura 5a. Tessiture originali.



Figura 5b. Tessiture sintetiche

Identificazione dei movimenti di ripresa [14]

Uno degli aspetti che caratterizzano sequenze di immagini, e che possono essere utilmente impiegate per operazioni di browsing, di editing assistito, etc. è l'insieme dei movimenti del sistema di ripresa delle immagini, detti anche *egomotion* (pan, tilt, zoom, rotazione). E' stata ulteriormente sviluppata una tecnica di stima dei parametri di egomotion.

Restauro di blocchi mancanti [4]

E' stata ulteriormente sperimentata una tecnica, già oggetto di studio, per la ricostituzione di parti mancanti di immagine mediante l'uso dei filtri armonici circolari. La tecnica consiste nello stimare mediante tali filtri le direzioni di strutture lineari (unidimensionali) esistenti attorno ai blocchi mancanti. Attraverso criteri di priorità compositiva, tali strutture vengono proseguite nella parte mancante mediante interpolazione pesata .

Segmentazione

Lo schema di analisi-sintesi di fig. 1 è stato utilizzato per un procedimento di segmentazione di immagine basato sull'uso simultaneo delle informazioni di movimento, colore e bordi. Un esempio dei risultati ottenuti è mostrato in fig. 6



Figura 6a. Immagine originale



Figura 6b. Immagine segmentata

Riferimenti bibliografici

- [1] S. Mallat, "Wavelets for a Vision" *IEEE Proceedings of IEEE*, Vol.84, NO.4, April 1996.
- [2] L. Capodiferro, S. Ventura, G. Jacovitti: "Enhanced Resolution Control for Video Sequences" *Wavelet Applications in Signal and image Processing*, San Diego, July, 1997.
- [3] L. Capodiferro, S. Puledda, G. Jacovitti: "Missing Block Recovery by Linear Pattern Propagation" *Picture Coding Symposium*, Berlino, Settembre 1997.
- [4] C. Becchetti, P. Campisi: "Autoregressive modelling of Texture using Sparse Supports and Bussgang Techniques", *DSP 1997*, Santorini, July 1997.
- [5] G. Jacovitti, A. Neri: "Circular Harmonic Wavelets: a Tool for Optimum Scale-Orientation Independent Pattern Recognition", *Wavelet Applications V, Proceedings of SPIE* Vol.3391, 1998.
- [6] G. Jacovitti, A. Neri: "Content based Image Classification with Circular Harmonic Wavelets", *Hybrid Image and Signal Processing VI, Proceedings of SPIE* Vol.3389, 1998.
- [7] P. Campisi, G. Jacovitti, A. Neri: "Optimal Wold Like decomposition of 2D random Fields" *EUSIPCO 98*, Rhodes, 8-11 September 1998.
- [8] G. Jacovitti, A. Neri: "Motion Field Estimation based on Laguerre-Gauss Circular Harmonic Pyramids", *Wavelet Applications VI, Proceedings of SPIE* Vol.3723, SPIE's Aerosense '99, 1999.
- [9] M. Carli, G. Jacovitti, A. Neri: "Riconoscimento di oggetti invariante rispetto alle rotazioni basato sull'espansione di Laguerre-Gauss" (vedi nella presente raccolta).
- [10] L. Capodiferro, G. Andreani, S. Puledda, G. Jacovitti: "Decomposition of Still and Video Images with Edge Segments" *Wavelet Applications in Signal and image Processing, Denver*, July, 1999.
- [11] P. Campisi, A. Neri, G. Scarano, "A Single Input Three Output Model Based Approach to Color Texture Generation", *IEEE International Conference on Image Processing 1998 (ICIP 98)*, pp.67-71, Chigago, Illinois (USA), 4-7 October 1998
- [12] P. Campisi, A. Neri, G. Scarano, "Reduced complexity modeling and reproduction of colored textures", *IEEE Transactions on Image Processing*, vol.47, no.2, February 2000 (sottomesso 1998).
- [13] G.Giunta, U. Mascia, "Estimation of Global Parameters by Complex Linear Regression" *IEEE Trans. on Image Processing*, Vol.8,NO.11, November 1999 (sottomesso 1998).