



Scuola Dottorale di Ingegneria Sezione di
Ingegneria dell'Elettronica Biomedica,
dell'Elettromagnetismo e delle
Telecomunicazioni
XXV Ciclo

**Tecnologie dell'Informazione e della
Comunicazione per la
valorizzazione dei Beni culturali**

Candidato

Paolo Francesco Sità
Matricola 11743-263

Docente Guida

Prof. Alessandro Neri

Coordinatore

Prof. Alessandro Salvini

Tesi presentata in parziale adempimento dei requisiti necessari
per conseguire il titolo di Dottore di Ricerca in Ingegneria

Aprile 2014

Tesi discussa il 15 Maggio 2014
di fronte a una commissione esaminatrice composta da:

Prof. Guglielmo D’Inzeo

Prof. Filiberto Bilotti

Prof. Franco Mazzenga

**Tecnologie dell’Informazione e della Comunicazione per la valorizzazione
dei Beni culturali**

Tesi di dottorato. Università di Roma - ROMATRE

© 2014 Paolo Francesco Sità. Tutti i diritti riservati

Questa tesi è stata composta con L^AT_EX .

Versione: 28 aprile 2014

*Dedicato alla
mia Famiglia*

Non sono i frutti della ricerca scientifica che elevano un uomo ed arricchiscono la sua natura, ma la necessità di capire e il lavoro intellettuale. (Albert Einstein)

Vorrei ringraziare il Prof. Alessandro Neri per avermi trasmesso parte della sua conoscenza e professionalità.

Un ringraziamento particolare all'Ing Licia Capodiferro con la quale ho condiviso momenti importanti della mia vita sia professionali che umani.

Un grazie alla Fondazione Ugo Bordoni e all'Università Roma Tre che hanno permesso la realizzazione di questo lavoro di dottorato.

Un grazie di cuore agli amici che mi hanno supportato e condiviso con me successi e fallimenti.

Grazie Luca, Raffaele, Andrea, Guido, Elena, Emiliano, Federica

Un grazie speciale a Simona.

Un doveroso e sentito ringraziamento alla mia famiglia che ha sempre creduto in me, donandomi preziosi consigli e riempiendomi di affetto.

Sommario

Il progetto di ricerca, svolto grazie alla collaborazione con la Fondazione Ugo Bordoni, si propone di studiare e sviluppare algoritmi e applicazioni per la raccolta e l'elaborazione di immagini e video digitali destinati allo studio, alla salvaguardia e alla fruizione su larga scala di beni culturali. Il progetto di ricerca è stato strutturato in due filoni principali: dal punto di vista tecnologico è stato realizzato l'applicativo software CHIP (Cultural Heritage Image Processing tool), uno strumento per la manipolazione e analisi di immagini relative ai beni culturali, mentre dal punto di vista della teoria sono state analizzate le tecniche di codifica alla base degli algoritmi sviluppati.

Esistono molti software commerciali per l'elaborazione delle immagini ma spesso tali sistemi sono concepiti per la manipolazione di immagini non strettamente correlate al campo del cultural heritage. Al fine di ottenere risultati migliori, in termini di qualità e prestazioni, è necessario utilizzare algoritmi specifici; l'applicativo CHIP nasce come strumento progettato e sviluppato con l'obiettivo di sfruttare le tecnologie digitali a favore della gestione, della valorizzazione e della tutela dei beni culturali.

All'interno della tesi verranno descritte le caratteristiche del sistema sviluppato, con particolare attenzione sull'analisi dei principali algoritmi sviluppati come il digital impainting, l'enhancement e l'High Dynamic Range. Verranno, inoltre, illustrate le tematiche affrontate durante la ricerca ovvero le principali tecniche legate alla rappresentazione tridimensionale di oggetti multimediali.

Il lavoro di ricerca ha portato alla realizzazione di un approccio innovativo per la valorizzazione e la fruizione dei beni culturali realizzando un sistema integrato multimodale e multimediale con flussi di navigazione semplici, intuitivi e veloci, con una interazione attiva, sociale e di qualità e con un target multilingue avente diversi livelli di lettura. In particolare nell'ambito del progetto Rome ArcheoMedia, svolto in collaborazione con la Fondazione Ugo Bordoni e la Soprintendenza Speciale per i Beni Archeologici di Roma è stato possibile, attraverso le nuove tecnologie multimediali, implementare un differente approccio alla comprensione dei beni archeologici nel loro contesto originale mediante percorsi di visita personalizzati e interattivi.

Le tecnologie utilizzate per la realizzazione di tour e ricostruzioni virtuali tridimensionali hanno permesso di concludere il lavoro di ricerca arricchendolo

di elementi multimediali ed aprendo nuove strade a future ricerche, fondendo nel miglior modo possibile la teoria con la pratica.



Indice

Elenco degli acronimi	vii
Sommario	x
1 Valorizzazione Beni Culturali	1
1.1 Introduzione	1
1.1.1 ICT per i beni culturali	2
1.2 Cultural Heritage	3
1.2.1 Gestione	5
1.2.2 Diagnosi e restauro	6
1.2.3 Tutela	8
1.2.4 Comunicazione, divulgazione e fruizione	9
2 Cultural Heritage Image Processing Tool	13
2.1 Introduzione	13
2.2 Enhancement	16
2.3 Algoritmo di compressione della dinamica	22
2.3.1 Tecnica omomorfica	22
2.3.2 Tecnica basata sulla compressione dei bordi	23
2.3.3 Algoritmo High Dynamic Range	25
2.4 Digital Impainting	33
2.5 Watermarking	42
2.5.1 Tecniche di watermarking	44
2.5.2 Dominio del watermarking	46
3 Structure From Motion	53
3.1 Introduzione	53
3.1.1 Fisiologia del sistema visivo umano	55

3.1.2	Modelli geometrici della fotocamera	58
3.1.3	Stereosopia	61
3.2	La Stereopsi	66
3.2.1	Principi base della Stereopsi	66
3.2.2	Calcolo della Stereopsi	71
4	Applicativi per la valorizzazione dei beni culturali	81
4.1	VisualSFM	81
4.2	Panotour	86
4.3	Aventino: applicazione Android	89
5	Conclusioni	91
A	Strumenti matematici	93
A.1	Espansione di Laguerre Gauss	93
A.2	Circular Fitting	102
A.2.1	Fitting Algebrico	102
A.2.2	Fitting Geometrico non lineare	103
A.2.3	Linear Least-Square	105
A.3	Prodotto di Kronecker	108
A.4	Decomposizione ai valori singolari	108
A.5	Prodotto esterno	109
A.6	Metodo DLT	109
A.7	Trasformata Discreta Coseno	110
A.8	Trasformata Discreta Wavelet	111
A.9	Algoritmo di Gauss-Seidel	114
	Indice delle Pubblicazioni	115
	Bibliografia	124

Capitolo 1

Valorizzazione Beni Culturali

1.1 Introduzione

L'attenzione progressiva che i Beni Culturali stanno ricevendo nella società contemporanea non solo da parte di tutti i generi di media: televisione, giornali, settimanali, libri, internet, ma anche e soprattutto da masse sempre più numerose di turisti, grazie alle straordinarie possibilità di mobilità oggi a disposizione, pone questo settore tra quelli di punta dell'economia. Ciò è tanto più vero per l'Italia, indiscusso paese-simbolo dei Beni Culturali. Sebbene a prima vista possa apparire contraddittorio, le moderne tecnologie, nelle loro diverse e molteplici espressioni, possono giocare un ruolo veramente cruciale nella valorizzazione del Patrimonio Culturale e, di riflesso, nella sua fruizione al fine di meglio motivarla, qualificarla e anche ampliarla; in definitiva, quindi, correlandola con l'intero indotto economico legato al turismo. La delicatezza propria del settore, tuttavia, richiede un ingresso misurato, discreto e sempre qualificato delle tecnologie.

Infatti lo sviluppo scientifico e tecnologico maturato negli ultimi decenni ha permesso una diffusione in larga scala della conoscenza e della cultura, rappresentati dalle tradizioni e dai costumi tramandati nel corso dei tempi, ovvero il patrimonio storico di una civiltà, di un paese, un bene che in quanto tale si ha l'obbligo morale di proteggere a memoria del nostro passato e nel rispetto delle generazioni future. Uno degli obiettivi principali delle tecnologie dell'informazione e della comunicazione è di favorire la conoscenza integrata del bene e del suo contesto affinché diventi un fattore di crescita culturale, sociale ed economica del sistema territoriale in cui si colloca. In tale prospettiva

abbiamo assistito ad una crescita esponenziale di nuove tecnologie in diversi campi di studio e di ricerca che hanno offerto utili strumenti per la tutela e la valorizzazione del patrimonio culturale del nostro paese. Le tecniche di digitalizzazione ed elaborazione di oggetti multimediali sono state applicate in maniera efficace e proficua alle procedure di catalogazione, restauro, analisi e tutela dei reperti storico artistici, garantendo, ad esempio, la creazione di immagini ad alta risoluzione di moltissimi dipinti, utili per la ricerca e l'integrità delle opere nel tempo. Inoltre non privo di interesse risulta il campo relativo alla protezione dei diritti di proprietà intellettuale e dei diritti di sfruttamento economico tutelati dalla legge, riguardo questo aspetto la digitalizzazione, attraverso tecniche di "digital watermarking", che permette una sicura ed efficiente gestione di tale processo.

Infine grande attenzione hanno suscitato le tecniche digitali stereo, 3D ed augmented reality per la fruizione e la gestione di manufatti appartenenti al patrimonio culturale.

1.1.1 ICT per i beni culturali

Il settore ICT (Information and Communication Technology) cioè Tecnologie dell'Informazione e della Comunicazione, è comunemente inteso come un insieme di tecnologie, che comprendono sia componenti hardware che software, e consentono la rappresentazione, l'organizzazione, l'elaborazione e lo scambio delle informazioni. Tali informazioni possono essere numeriche, testuali, visive, sonore o costituite da una combinazione totale o parziale di queste, in tal caso prenderanno il nome di contenuti multimediali. Tuttavia, questa interpretazione si limita a considerare i soli ambiti puramente tecnologici, dell'elettronica, dell'informatica e delle telecomunicazioni, senza considerare che le applicazioni dell'ICT investono ormai qualsiasi settore della società, con tutte le conseguenze economiche e organizzative sulle competenze, le professionalità, la formazione e la ricerca nell'intero settore.

Possiamo quindi affermare che l'ICT è uno dei temi più importanti, innovativi e quindi strategici per la politica, l'economia, l'imprenditoria, la ricerca e, in generale, tutto il mondo dei media.

Tra le innumerevoli innovative applicazioni, legate all'ICT, si può sicuramente annoverare la virtualizzazione 3D che in questi anni più recenti ha subito un impulso e una rilevanza sempre maggiore.

Tale tecnica ha permesso la possibilità di rappresentare in forma digitale tutta la realtà tridimensionale, sia statica che dinamica, in cui viviamo. Costituisce, infatti, la naturale evoluzione delle tecniche informatiche di digitalizzazione che, a partire dalla rappresentazione binaria dei numeri utilizzata fin dalla nascita degli elaboratori elettronici in quella che possiamo definire la preistoria dell'informatica, ha consentito la gestione digitale e virtuale di simboli, di testi, d'immagini bidimensionali, di suoni e poi ancora di filmati.

In questo ambito si può rilevare come queste applicazioni possano avere nei prossimi anni un enorme sviluppo tanto da rivoluzionare la nostra vita quotidiana.

Il processo di rappresentazione digitale della realtà prevede varie fasi e passaggi, ciascuno a suo modo innovativo e impegnativo: l'acquisizione ed elaborazione digitale di immagini e modelli tridimensionali, la ricostruzione tridimensionale e quindi la rappresentazione virtuale di oggetti, ambienti, strutture e il loro eventuale movimento.

In questa direzione, che è stata prima di ricerca e poi di applicazione, si va a inquadrare il tema della fruizione di siti di notevole interesse storico-archeologico e naturalistico, soprattutto quando essi sono difficilmente accessibili agli studiosi o al grande pubblico. Il fatto che il patrimonio dei Beni culturali e ambientali rappresenti una delle risorse del nostro paese, come riconosciuto anche a livello internazionale, offre una enorme opportunità di intervento di ulteriore sua valorizzazione per le possibili applicazioni delle metodologie e delle tecnologie della realtà virtuale. È ben evidente come tutto ciò possa rappresentare un fattore decisivo di crescita dell'intera economia del nostro paese.

1.2 Cultural Heritage

Con il termine beni culturali intendiamo l'insieme di artefatti oggetti, pitture, sculture e edifici, ma anche tradizioni, costumi, che in qualche modo costituiscono il nostro patrimonio culturale, tramandatoci dai nostri antenati. In Italia, appartengono a questa sfera parchi archeologici, musei, oggetti di valore storico-artistico, edifici rilevanti, monumenti, chiese, quadri, mobili, arredi, oggetti di culto, tessuti, ed anche città, strutture urbanistiche, tradizioni popolari, tradizioni alimentari, tradizioni culinarie e tradizioni religiose.

Non ci riferiamo alla natura fisica dell'oggetto, né al suo valore estetico, infatti gli oggetti inquadrabili, in senso stretto, come arte costituiscono solo una frazione dell'immenso del nostro patrimonio culturale.

Possiamo affermare che il termine culturale riferito ad oggetto d'interesse, indipendentemente dalla sua natura, racchiude in se molteplici aspetti, primo tra tutti la sua storia, scoprire se l'oggetto ne possiede una e studiare la sua importanza dal punto di vista sociale, economico e religioso, indagando, inoltre su la possibile correlazione con la società odierna, per stabilirne, quindi, il suo valore, non puramente economico.

Le tecnologie dell'informazione e della comunicazione rappresentano un valido strumento per garantire al settore dei Beni Culturali una serie di servizi indispensabili che potremmo sintetizzare in:

- **Gestione, studio e ricerca:** sono interessate le applicazioni che permettano una efficace gestione del patrimonio culturale, ottimizzando, lo stesso tempo i costi, e che favoriscono le attività di studio e di ricerca inerenti ai beni culturali, garantendo un facile accesso alle informazioni dislocate sul territorio.
- **Diagnosi restauro:** le applicazioni e gli strumenti che forniscono al personale addetto ai lavori di ottenere una diagnostica dello stato di conservazione e strumenti utili per l'attività di restauro attraverso sempre più sofisticati mezzi tecnologici.
- **Tutela:** le applicazioni che consentono di tutelare il patrimonio culturale, sia contro atti criminosi sia rispetto a calamità naturali.
- **Formazione:** le applicazioni rivolte ai giovani studenti per aiutarli negli studi, permettendo loro di avvicinarsi alla cultura in modo diverso e più consono alle loro attitudini.
- **Comunicazione, divulgazione e fruizione:** le applicazioni che sia permettono di comunicare al pubblico la rilevanza e la vastità del nostro patrimonio culturale, sia a scopo divulgativo generale che per favorire il turismo, sia garantire una fruizione più efficace e immediata al patrimonio culturale.

Per maggior chiarezza riguardo la precedente classificazione ci apprestiamo a stilare una più dettagliata descrizione dei singoli aspetti, soffermandoci sugli

argomenti che maggiormente hanno attirato la nostra attenzione e sviluppati all'interno del lavoro di ricerca.

1.2.1 Gestione

Gli organi predisposti alla gestione, ovvero alla conoscenza di quali sono i vari beni, le loro caratteristiche generali, la loro collocazione, il loro stato di manutenzione ed il loro effettivo valore, sono le sovrintendenze, le regioni, le province, i musei, le diocesi, le banche e tutti i privati proprietari di beni culturali rilevanti. A queste figure sono sostanzialmente destinate le applicazioni dedicate alla gestione del patrimonio culturale, che si traducono in applicazioni di basi di dati che permettano di inventariare tutti gli oggetti di rilevante interesse attraverso l'uso di opportune schede organizzate in sistemi, che seppur complessi dal punto di vista strutturale, devono essere facilmente accessibili. Questa operazione risulta per i beni culturali spesso più complessa di quanto sembri, infatti la mole di oggetti presenti, per esempio, in un'area archeologica può sfiorare le diverse migliaia o decine di migliaia, da un tempio ad un frammento d'anfora.

È necessario, quindi, sviluppare un metodo efficace ed adattabile a tutte le situazioni che permetta una facile catalogazione con un sistema di ricerca sviluppato su più criteri.

L'obiettivo principale di una gestione efficace del patrimonio culturale giace nella realizzazione del magazzino digitale, che permette di sapere quello che ci dovrebbe essere e avere la possibilità di verificarne la presenza effettiva. Questo rappresenta un utile supporto all'attività di studio e ricerca, garantendo al ricercatore di consultare, ovunque esso si trovi, qualunque scheda presente nell'archivio e al tempo stesso conoscere l'ubicazione dell'oggetto di interesse. Inoltre la catalogazione permetterebbe la pianificazione delle azioni di tutela, o del censimento delle opere in caso di furto o alluvione, garantendo l'ulteriore possibilità di poter dare un valore al patrimonio, dando un potenziale valore economico a ciascun oggetto inventariato, e stimando così il valore potenziale del nostro patrimonio culturale.

Purtroppo la maggior parte del patrimonio culturale italiano non è inventariato, o inventariato male, con modelli di rappresentazione dei dati antiquati, sebbene negli ultimi anni è stato fatto uno sforzo notevole per la realizzazione di un archivio universale, primi tra tutti il MiBACT (Ministero dei beni e

delle attività culturali e del turismo) che, per citarne alcuni, ha sviluppato due progetti europei per la divulgazione e catalogazione dei beni culturali: Minerva (<http://www.minervaeurope.org/>) e Michael(<http://www.michael-culture.org/about.html>), siamo ben lontani dal raggiungimento dell'obiettivo, infatti la presenza di problemi burocratici e la scarsa comunicazione tra la miriade di enti esistenti rappresenta un grosso ostacolo da superare.

Comunque uno strumento fornito dall'ICT, man mano sempre più diffuso, è la georeferenziazione dei siti culturali, che ha permesso la realizzazione di una mappa nella quale è possibile trovare l'ubicazione esatta del sito di interesse e tutte le informazioni necessarie richieste dall'utente.

1.2.2 Diagnosi e restauro

La diagnostica applicata ai beni culturali ha lo scopo di ricavare informazioni dettagliate e accurate sullo stato di un bene, sia esso un vaso, una scultura, un affresco o un edificio, e, se possibile, senza procurare danni al bene stesso. Qualsiasi prelievo fisico in generale, è da escludersi.

In particolare, l'applicazione della microscopia ottica ed elettronica nel campo dei Beni Culturali fornisce un contributo scientifico di estremo interesse. Infatti, è applicata per indagini diagnostiche finalizzate sia al restauro, individuazione di tecniche artistiche, analisi della componente pittorica, analisi di materiale artistico di varia natura, ecc, sia alla conservazione di opere di notevole prestigio storico-artistico.

Le tecniche utilizzate, in relazione alla natura del bene e del danno da rilevare, si possono basare sui **raggi-X**, per rilevare, ad esempio, la natura dei vari materiali attraverso l'utilizzo di una microsonda elettronica che è sfrutta una tecnica di microanalisi elementale semiquantitativa della superficie dei materiali, non distruttiva per il campione, e permette di risalire alla composizione chimica, mettendo in evidenza particolari relativi alla tecnica pittorica o fenomeni di alterazione e di degrado presenti nell'opera d'arte. In particolare la microsonda elettronica (EDS), utilizzando le proprietà dei raggi X emessi dalla superficie del campione colpita dagli elettroni, permette di eseguire analisi chimiche del campione, fornendo informazioni dettagliate sulla natura chimica delle sostanze che li compongono. L'utilizzo di **raggi laser** permette di rilevare, per esempio, i diversi andamenti di una superficie, mediate l'utilizzo di complesse apparecchiature laser, in grado di operare in modo sistematico nella velocità

di acquisizione, si arriva a migliaia di punti al secondo, e nella possibilità di accedere direttamente ai dati in tempo reale, che forniscono una scansione tridimensionale della superficie in esame.

Altrettanto utile si è rivelato l'impiego di un sistema basato sulle **onde sonore** che permette di intercettare le diverse vibrazioni emesse, in reazione a uno stimolo sonoro, da un oggetto sia superficiali sia interne. Per emissione acustica si intende il rilascio di energia elastica da parte di un materiale ogni volta che si verificano nel suo interno modificazioni di carattere strutturale. Tale rilascio è dovuto alle microinstabilità locali che si producono nel corso dell'evoluzione di una struttura verso stati a diverso contenuto entropico. In questa situazione si generano movimenti dinamici locali, con conseguente formazione di onde elastiche di pressione di tipo vibrazionale che si propagano nel materiale fino a raggiungere la sua superficie. In tal modo le onde vengono trasmesse all'esterno e possono così essere rilevate mediante opportuni sensori. Storicamente la più antica applicazione dell'emissione acustica è nella sismologia, ma in seguito l'impiego della tecnologia è stata adoperata come mezzo analitico per lo studio delle caratteristiche del comportamento meccanico-strutturale dei materiali. In particolare la tecnica di esame prevede la disposizione di un certo numero di sonde piezoelettriche sulla superficie del materiale da esaminare per rilevare così l'energia acustica emessa spontaneamente, localizzando le coordinate delle sorgenti mediante tecniche di triangolazione. La strumentazione consiste in un analizzatore elettronico computerizzato dei parametri dello spettro dei segnali, i cui valori di insieme permettono di caratterizzare e riconoscere le varie sorgenti. I parametri più utilizzati sono: il numero dei treni d'onda e delle oscillazioni, l'energia associata, la durata, l'ampiezza massima. La tecnica, estremamente sensibile ai difetti, viene utilmente impiegata in tutti i settori industriali e della ricerca, ma è utilizzata per la risoluzione ad importanti problematiche strutturali nel campo dei beni culturali.

Le tecnologie informatiche e delle telecomunicazioni, in genere, hanno un ruolo da protagoniste in queste attività di diagnostica, infatti servono, innanzitutto, a filtrare in vario modo il segnale per aumentarne la leggibilità, a memorizzare in una base di dati i vari rilievi effettuati, ed inoltre permettono analisi non invasive attraverso l'uso di fotocamere ad alta risoluzione che restituiscono ricostruzioni tridimensionali dell'intera superficie dell'oggetto, anche di opere d'arte fragili o deformate, caratterizzate da geometrie superficiali molto com-

plesse.

È bene dire che la precisione con la quale è effettuata la rilevazione risulta inferiore alle tecniche elencate in precedenza, ma i risultati si dimostrano accettabili, ed inoltre grazie alle dimensioni delle apparecchiature, spesso ridotte, e al loro costo sicuramente occupano un posto di rilievo nella fase di diagnosi.

Parlando adesso di restauro risulta evidente che è necessario modificare lo stato di un bene culturale, o ai fini di una migliore conservazione, o per ripristinare uno stato che in qualche modo si ritiene più corretto. Come per la diagnosi, anche in questo caso, le tecnologie dell'informazione sono di grande utilità per la gestione dei dati e l'accesso, anche remoto, agli stessi.

Informatica, in particolare la computer vision e le telecomunicazioni danno il loro contributo alle tecniche di restauro. L'enorme vantaggio è rappresentato dalla possibilità di creare simulazione virtuali del restauro, come verrà dimostrato nel corso del lavoro di ricerca svolto. Infatti è stato possibile realizzare una tecnica automatica o semiautomatica che permette di assistere gli addetti ai lavori durante la fase del restauro stesso, monitorando costantemente i singoli step del lavoro.

1.2.3 Tutela

La tutela del patrimonio dei Beni Culturali in Italia è, naturalmente, una priorità, data la ricchezza del patrimonio stesso, ma si tratta, nello stesso tempo di un'attività difficile, data l'estrema frammentazione del patrimonio e la sua distribuzione sul territorio. Basti pensare al fatto che il patrimonio dei Beni Culturali, nel nostro Paese, si trova non solo i luoghi istituzionali come musei, palazzi presidiati e parchi archeologici nazionali, ma anche presso una miriade di realtà poco o per nulla presidiate come chiese, palazzi privati e piccole collezioni.

Per quanto riguarda la tutela è necessario distinguere due filoni principali.

La tutela contro le calamità naturali, che sfruttando le potenzialità offerte dall'informatica e dall'ICT, è realizzata utilizzando mappa che tiene conto della presenza sul territorio di tutti beni culturali, una sorta di base di dati che cataloghi tutti i beni a rischio potenziale, con le informazioni necessarie per decidere cosa fare per un pronto intervento. Anche se non completamente realizzata tale soluzione prende sempre più piede a livello nazionale. Il secondo

importante aspetto della tutela riguarda la prevenzione e la repressione del crimine. Il patrimonio dei beni culturali è così diffuso, che una sua protezione efficace è veramente difficile. I carabinieri hanno creato da tempo un nucleo specializzato, per la tutela del patrimonio culturale, molto efficace, ma sprovvisto di una infrastruttura sul territorio adeguata. Il ruolo delle ICT, in questo caso, potrebbe essere quello di fornire una base di dati distribuita, accessibile facilmente da tutto il territorio nazionale, ma estremamente protetta. La Banca Dati dei beni culturali illecitamente sottratti, la cui gestione è assegnata all'Ufficio Comando - Sezione Elaborazione Dati, è un elemento caratterizzante del Comando Carabinieri T.P.C. costituisce uno strumento unico nella sua concezione, per flessibilità degli applicativi, quantità dei dati trattati e capacità della risposta operativa. La sua origine, risalente al 1980, la qualifica come prima banca dati costituita nello specifico settore, che è tuttora unanimemente riconosciuta essere il database dedicato più ampio esistente al mondo. In essa sono quotidianamente inserite tutte le informazioni descrittive e fotografiche relative ai beni culturali da ricercare, che pervengono dai Reparti territoriali dell'Arma, dalle altre Forze di Polizia, dalle Soprintendenze, dagli Uffici Doganali e da Interpol per i beni all'estero. La Banca Dati, quindi, proprio in ragione dell'utilizzo di una sofisticata tecnologia informatica e delle numerose informazioni in essa contenute, è ormai divenuto uno strumento di avanguardia per l'elaborazione e l'analisi dei fenomeni criminali concernenti i beni culturali, in grado di indirizzare l'attività preventiva e investigativa dei vari reparti. Aggiornata giornalmente, è strutturata in moduli che consentono sia l'inserimento e la ricerca di eventi, persone, oggetti e loro possibili relazioni. Presenta un'interfaccia web con supporto multilingua e permette modalità di ricerca visuale e capacità di georeferenziazione di eventi, con l'interazione in tempo reale con palmari e personal computer portatili, agevolando la redazione di documentazione sul luogo dell'intervento e la consultazione e l'alimentazione dirette.

1.2.4 Comunicazione, divulgazione e fruizione

La comunicazione, applicata ai beni culturali, è uno dei temi più importanti in Italia. Possiamo affermare che il nostro paese ricco dal punto di vista culturale e poiché cultura vuol dire storia che si può narrare a partire da un bene culturale, allora comunicazione e cultura diventano quasi la stessa cosa. La cultura non è

legata alle caratteristiche fisiche degli oggetti, ma alla loro capacità di suscitare emozioni, risvegliare curiosità, quindi comunicare equivale a fare cultura e fare cultura equivale a comunicare. La valorizzazione dei beni culturali permette di rendere importanti i beni culturali agli occhi della società con il conseguente ritorno di risorse economiche verso il mondo dei beni culturali, provenienti sia dalla pubblica amministrazione che da sponsor privati.

È bene sottolineare che quando parliamo di valorizzazione non intendiamo meramente monetizzare ma piuttosto aumentare le risorse economiche, in seguito ad una corretta comunicazione. Infatti una adeguata forma di comunicazione porterebbe le istituzioni culturali ad essere popolari e presenti nella società in modo attivo e vivace attraendo grandi risorse economiche e fornendo servizi a costi molto contenuti.

Adesso forniamo una panoramica sulle applicazioni multimediali fornite dalle dal settore dell'ICT che si afferma leader nel settore e mette le radici prepotentemente nella comunicazione, divulgazione e fruizione di contenuti multimediali nel settore dei beni culturali.

Innanzitutto citiamo le applicazioni multimediali interattive disponibili principalmente su dispositivi fisici, come ad esempio i DVD, che rendono accessibile il patrimonio culturale. Ormai largamente superati ebbero grande successo negli anni '80 e '90, ma sono state in gran parte archiviate e sostituite dal Web, ormai alla portata di tutti. Attualmente tali applicazioni si trovano, sempre più raramente nelle scuole e, ancor di meno, in prodotti editoriali. Inoltre, poiché la qualità dei siti Web ha pressoché raggiunto la qualità DVD e quest'ultimi richiedono una catena distributiva per renderle accessibili al pubblico, tali applicazioni sono destinate a scomparire.

Per quanto riguarda la diffusione di materiale relativo al patrimonio culturale siti web senz'altro detengono un posto di rilievo e rappresentano, quindi, il maggior veicolo per la diffusione di applicazioni multimediali interattive nell'ambito dei beni culturali. Grazie alla loro proprietà di fornire accesso immediato ai contenuti da tutto il mondo e alla loro facilità di realizzazione ne ha favorito una diffusione a macchia d'olio. L'evoluzione dei siti web è stata rapida e sempre più interattiva, si è partiti da sterili siti che presentavano l'istituzione, ad esempio un museo, fornendo poche informazioni riguardo la natura stesse dei reperti contenuti, passando, in seguito, a siti che proponevano solo una dettagliata alle collezioni permanenti, fino ad arrivare, ai nostri giorni, a veri

e propri portali di informazione contenenti descrizione delle mostre ed eventi temporanei, mostre virtuali, giochi educativi e applicazioni per le scuole.

Quindi grazie al continuo progresso delle tecniche di comunicazione le istituzioni hanno la possibilità di raggiungere un ampio pubblico e a svolgere il loro ruolo di diffusori di cultura. Infine possiamo affermare che internet, grazie alla sua straordinaria evoluzione, oggi, ha reso possibile l'accesso a tutti i maggiori musei del mondo, ma anche, e soprattutto, i piccoli musei, che rappresentano una presenza attiva e stimolante, meritandosi l'appellativo di straordinario museo virtuale.

Per quanto riguarda la fruizione di contenuti legati strettamente ai beni culturali un ruolo decisivo è in corso di svolgimento da parte della tecnica della realtà aumentata, in lingua anglosassone *Augmented Reality (AR)*. L'AR è tecnologia che permette di sovrapporre alla realtà vista su uno schermo o attraverso uno schermo, uno strato di realtà virtuale. Al momento possiamo distinguere due grandi filoni di applicazione, uno sul pc e uno sul cellulare. Sul PC è mostrato alla webcam un oggetto, un simbolo grafico per attivare un programma che mostra una immagine tridimensionale sullo schermo. Sul cellulare, in genere uno smartphone, cosa molto più interessante, è possibile tralasciare la realtà circostante attraverso lo schermo, e vedere informazioni, dati, indicazioni sovrimpresse alla realtà. Il sistema di realtà aumentata permette, ad esempio, di ricollocare virtualmente monumenti e siti archeologici nel contesto d'origine, nello stesso momento in cui il visitatore sta effettuando la visita del medesimo sito o monumento in presenza e nella realtà. Tale sovrapposizione e penetrazione del reale con la ricostruzione virtuale potrebbe avvenire o all'interno di occhiali speciali che i visitatori indossano durante la visita o sul proprio telefono cellulare. Il sistema offre, perciò, una vera e propria guida personalizzata e gestibile simultaneamente alla visione reale del monumento e alla sua ricollocazione nella ricostruzione virtuale del suo contesto originario, infatti, il sistema è in grado di allinearsi automaticamente alla posizione del visitatore rispetto ai monumenti, essendo in esso integrato un sistema di orientamento, come un GPS. La realtà aumentata crea una nuova modalità di visita e di conoscenza dei beni architettonici, archeologici e paesaggistici, poiché, da passivo spettatore, il visitatore diventa con questo strumento, il vero protagonista attivo dell'azione culturale.

La visita aumentata si arricchisce sia della componente di ricostruzione 3D sia

di elementi conoscitivi e informativi, assecondando in tempo reale le richieste del visitatore.

Un esempio è rappresentato dalle Domus Romane di Palazzo Valentini, museo situato in Roma. Il progetto curato da storici dell'arte, architetti e architetti dell'amministrazione provinciale ha permesso di realizzare un suggestivo percorso tra i resti delle domus patrizie di età imperiale. La visita multimediale, che si avvale della tecnologia Spatial Augmented Reality (SAR) nella quale la realtà virtuale è direttamente proiettata sulle opere architettoniche e fruibile senza l'ausilio di nessun dispositivo, permette al visitatore, grazie ad un gioco di luci, suoni e ricostruzioni virtuali di essere totalmente immerso in un viaggio reale-virtuale durante il quale vedrà rinascere ambienti decorazioni e arredi.

Un'ultima applicazione degna di nota nell'ambito della fruizione dei beni culturali è la ricostruzione virtuale. Essa rappresenta la riproduzione virtuale del mondo reale, interamente realizzata attraverso strumenti, sempre più avanzati, di computer grafica 3D. Lo spettatore attraverso un dispositivo di visualizzazione, come un monitor e un dispositivo di input, ad esempio un pad di controllo, interagisce con il mondo virtuale, come nel caso dei videogiochi. L'utente può esplorare l'ambiente virtuale ed interagire con eventuali oggetti presenti nella scena, accompagnato da suoni che rispecchiano la realtà. Citiamo, come esempio, il Museo Archeologico Virtuale dell'Iraq, che riproduce virtualmente i contenuti originali del museo di Bagdad. Il progetto, a cura del Ministero degli Affari Esteri e il Consiglio Nazionale delle Ricerche, rappresenta un valido strumento di apprendimento interattivo corredato da una ottima ricostruzione 3D, e ricco di reperti virtuali, filmati, elaborazioni cartografiche di siti archeologici.

Capitolo 2

Cultural Heritage Image Processing Tool

2.1 Introduzione

L'applicativo software CHIP sviluppato durante il lavoro di ricerca si è ispirato ai numerosi strumenti dedicati all'elaborazione delle immagini presenti sul mercato, prestando particolare attenzione all'ambiente dei beni culturali. I punti di forza dell'applicativo software, in continuo aggiornamento, sono l'interfaccia semplice ed usabile, ed i tempi di calcolo, necessari per l'esecuzione dei comandi, che risultano ragionevolmente brevi.

Il cuore del sistema è rappresentato da una libreria software, sviluppata sfruttando le potenzialità del linguaggio di programmazione C++, ricca di funzioni e algoritmi per l'elaborazione delle immagini, già noti in letteratura o progettati durante lo sviluppo del lavoro di ricerca. Le funzionalità, che vanno dal semplice calcolo dell'istogramma alla ben più complessa ricerca di punti salienti presenti nelle immagini, lavorano sinergicamente per garantire un sistema performante dal punto di vista computazionale e qualitativo.

L'interfaccia grafica, sviluppata secondo i principi dell'ergonomia cognitiva, integra nel miglior modo possibile, le capacità e le esigenze dell'uomo con le caratteristiche della macchina. Per completezza ricordiamo la definizione di ergonomia cognitiva presente nello statuto della Società Europea di Ergonomia Cognitiva (EACE, costituita nel 1987), ovvero: l'ergonomia cognitiva ha come oggetto di studio l'interazione tra il sistema cognitivo umano e gli strumenti per l'elaborazione di informazione. La conoscenza prodotta da questo studio è

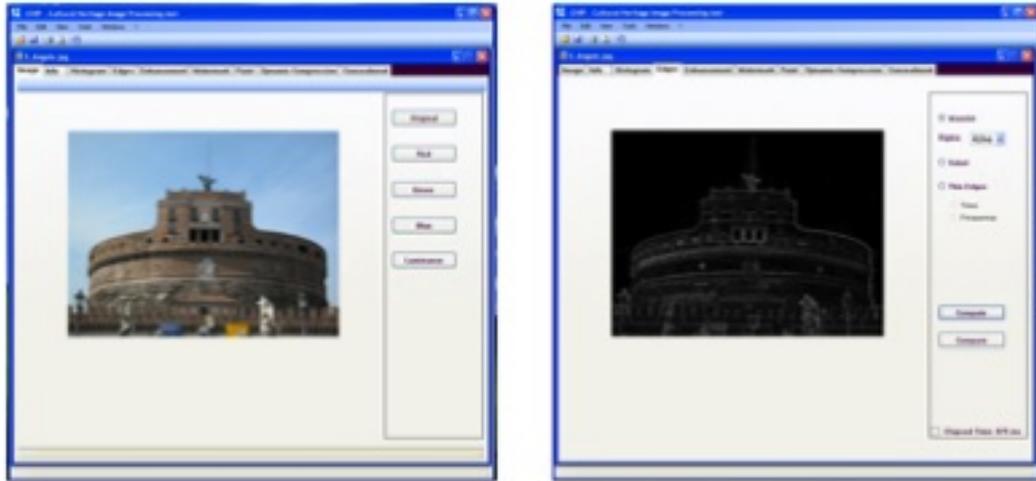


Figura 2.1. Tab principale (sinistra) e tab estrazione edge (destra)

utilizzata per supportare la progettazione di strumenti appropriati per i più svariati usi, dal lavoro, all'educazione, al divertimento. Sulla base di questo, utilizzando il linguaggio C#, l'interfaccia si presenta sobria ed essenziale, nella parte superiore vi sono dei menù contestuali a tendina attraverso i quali è possibile caricare, elaborare e salvare le immagini. L'organizzazione delle finestre è organizzata in tab, ognuno dei quali fa riferimento ad una particolare funzione di elaborazione.

Utilizzando il menù carichiamo l'immagine da elaborare, verrà visualizzata nel tab principale (figura 2.1), dove è possibile, adoperando gli appositi pulsanti, estrarre le singole componenti RGB e luminanza.

Passando al tab "Edge" (figura 2.1), possiamo scegliere uno tra gli algoritmi proposti per l'estrazione e la visualizzazione dei bordi presenti nell'immagine. Accedendo ai tab "Info" e "Histogram" (figura 2.2) visualizzeremo rispettivamente le informazioni relative ai metadati incapsulati nell'immagine e la distribuzione statistica di valori delle componenti di colore RGB e della luminanza dell'immagine.

Le altre funzionalità presenti nell'applicativo, descritte minuziosamente nei paragrafi successivi, sono:

- **Enhancement**, algoritmo utilizzato per il miglioramento della qualità visiva delle immagini affette da problemi di sfocatura
- **High Dynamic Range**, algoritmo sviluppato con lo scopo di ottenere una corretta esposizione di tutte le porzioni delle immagini.

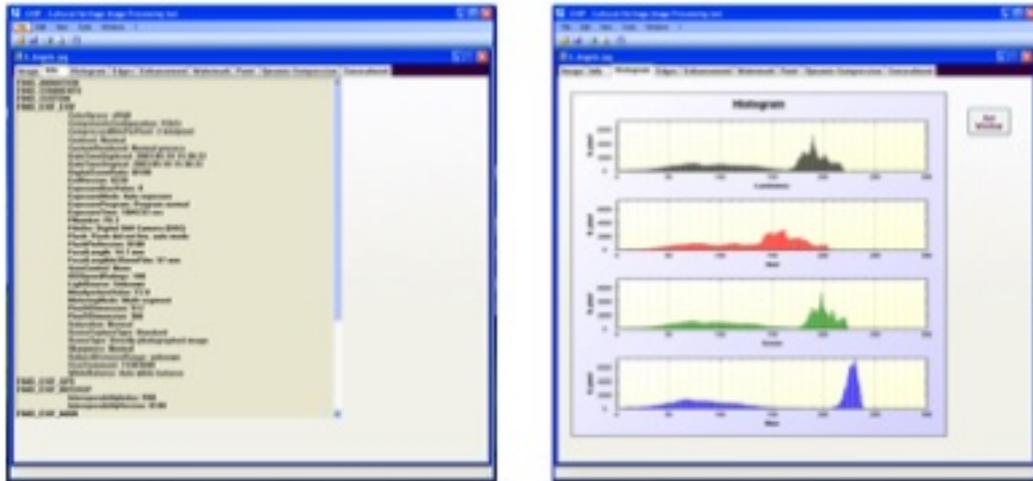


Figura 2.2. Tab Info (sinistra) e tab Histogram (destra)

- **Digital Impainting:** tale algoritmo è un restauro digitale dell'immagine e permette di ricostruire parti mancanti dell'immagine, garantendo uniformità e coerenza con il contesto stesso dell'immagine.

2.2 Enhancement

Viene proposta un algoritmo per il miglioramento visuale di immagini digitali defocalizzate e, in generale, per il miglioramento dell'aspetto complessivo di immagini di non elevata qualità, sia in termini di nitidezza che di contrasto. La procedura, applicabile sia ad immagini monocromatiche, rappresentate mediante scala di grigi, che a colori, si basa sulla decomposizione a multirisoluzione mediante wavelet circolari armoniche di Laguerre – Gauss, sull'elaborazione delle sue componenti e sulla ricostruzione dell'immagine restaurata utilizzando i piani elaborati. Tipiche applicazioni della procedura si hanno in tutti i campi della scienza e della tecnica in cui le difficili condizioni operative pregiudichino la possibilità di ottenere immagini di buona qualità. In ambito scientifico possiamo pensare ad immagini mediche con finalità diagnostiche o ad immagini aeree o da satellite per la topografia o il controllo del territorio. In campo artistico e commerciale possiamo citare il restauro di immagini fotografiche di valore storico o culturale o il restauro di pellicole cinematografiche o sequenze video, anche contestualmente alla loro memorizzazione sui nuovi supporti digitali oggi disponibili. La sequenza di operazione da seguire per ottenere il risultato finale può essere riassunta nei passi principali:

- acquisizione immagine
- estrazione piani wavelet immagine
- attenuazione piano a bassa frequenza
- elaborazione piani wavelet ad alta frequenza
- ricostruzione immagine

L'immagine in ingresso è rappresentata secondo lo standard RGB, ad ogni pixel è associata una terna di numeri compresi tra 0 e 255. Per evitare problemi dovuti alla non linearità di alcune operazioni effettuate è stato scelto di convertire l'immagine nel formato YUV, mediante una opportuna matrice di trasformazione.

$$\begin{bmatrix} Y_{601} \\ U_{601} \\ V_{601} \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.500 & -0.419 & -0.081 \\ -0.169 & -0.331 & 0.500 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2.1)$$

L'algoritmo elabora la componente di luminanza Y e lascia inalterate le componenti di cromaticanza U e V , utilizzate solo in fase di ricostruzione.

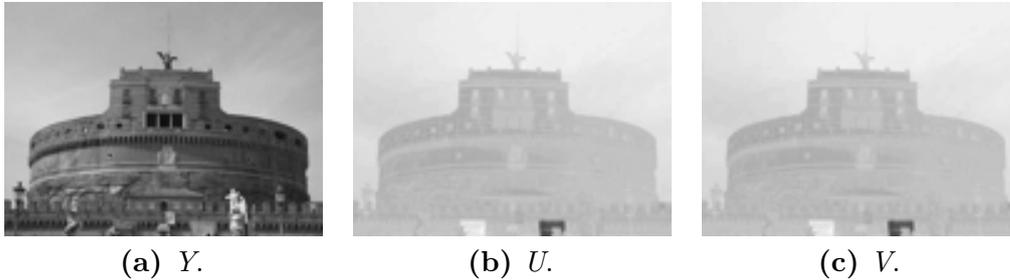


Figura 2.3. Componenti: Luminanza Y , Crominanza U e Crominanza V

La fase successiva è rappresentata dall'estrazione dei piani wavelet, in tale dominio, quello della frequenza, la rappresentate dell'immagine viene ricavata mediante filtri armonici di Laguerre-Gauss, calcolando la convoluzione con la componente di luminanza dell'immagine da elaborare.

I filtri utilizzati sono quelli con $k = 0, n = 0, 1, 2, 3, 4$ alle risoluzioni $s_j = 1, 2, 4$. Si ottengono 12 piani ad alta frequenza ed un piano a bassa frequenza ($n = 0, s_j = 1$).

Poiché l'operazione di sfocatura di un'immagine, nel dominio della frequenza, è rappresentata attraverso un'attenuazione della parte alta dello spettro, allora un'immagine sfocata può essere ottenuta andando ad attenuare le alte frequenze della trasformata di Fourier dell'originale. Per giustificare tale proprietà basti notare che un bordo ideale, ossia una variazione a gradino della luminanza, ha una trasformata di Fourier che copre tutto lo spettro disponibile, mentre una variazione graduale della luminanza tra un valore basso ed uno alto, ossia un bordo sfocato, ha una trasformata di Fourier collocata alle basse frequenze. Consideriamo ora il piano wavelet a bassa frequenza, lo attenuiamo con un coefficiente lineare costante β_{LF} minore di 1, utilizzando la formula

$$LF'[l][m] = \beta_{LF} \cdot LF[l][m] \quad (2.2)$$

allo scopo di ripristinare l'equilibrio tra le basse e le altre frequenze dello spettro e quindi recuperare la qualità visiva dell'immagine stessa. Poiché attenuando il piano wavelet a bassa frequenza, contenente informazioni sulle zone uniformi dell'immagine, sono alterati sia il contenuto spettrale alle basse frequenze, sia il valore medio dell'immagine, ossia la luminanza dei pixel

nel dominio dello spazio. Per rimediare a questo effetto indesiderato, che provoca uno scurimento dell'immagine restaurata, che quindi risulta meno luminosa dell'originale, è necessario agire sull'immagine ricostruita ripristinando il corretto valore di luminanza dei pixel. Quindi moltiplichiamo il valore di ogni pixel con un opportuno coefficiente che, data la linearità dell'elaborazione effettuata sul piano LF, è semplicemente il reciproco del coefficiente utilizzato per l'attenuazione.

Per quanto riguarda l'elaborazione dei piani ad alta frequenza è necessario fare alcune considerazioni. In particolare il modulo dei piani wavelet di ordine 1 è proporzionale alla variazione locale di luminanza, mentre la fase indica l'orientamento spaziale dei bordi. Inoltre passando da un'immagine originale alla sua versione sfocata, le tracce dei bordi sui piani wavelet di ordine 1 si allarghino e diminuiscano di intensità.

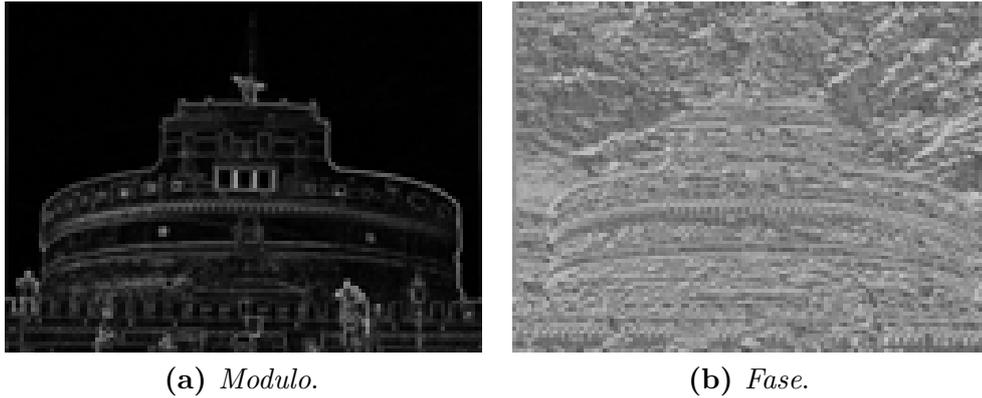


Figura 2.4. Modulo e fase del piano wavelet con indici $n=1$, $k=0$ al alta risoluzione

L'operazione di sfocatura provoca una inevitabile alterazione dei bordi che si ripercuote sui piani wavelet. Allo scopo di contrastare l'effetto introdotto dall'operazione di sfocatura il nostro metodo genera il piano restaurato, elaborando i piani di ordine 1, mediante la formula quadratica:

$$|HF'_{1,j}[l][m]| = \frac{\alpha_{1,j}}{Max_{loc}} \cdot |HF_{1,j}[l][m]|^2 + \beta_{1,j} \cdot |HF_{1,j}[l][m]| \quad (2.3)$$

Tale piano prodotto è molto utile per identificare quei bordi che figurano a tutte le risoluzioni. I coefficienti $\alpha_{1,j}$ e $\beta_{1,j}$ sono scelti in base al grado di sfocatura, mentre Max_{loc} è il massimo locale calcolato in un opportuno intorno del pixel considerato.



Figura 2.5. Maschera di ordine 1

La formula non viene applicata a tutti i punti dell'immagine, ma solo a quelli che soddisfano i vincoli imposti dalla maschera M_1 ottenuta nel seguente modo:

$$M_1[l][m] = \begin{cases} 1 & \text{se } PP_1[l][m] > \gamma_1 \\ 0 & \text{altrove} \end{cases} \quad (2.4)$$

dove γ_1 è una soglia scelta opportunamente, mentre $PP_1[l][m] = \prod_{j=0}^2 |HF_{1,j}[l][m]|$ è il piano prodotto, ottenuto moltiplicando i moduli dei piani wavelet, tutti di ordine 1, ad alta, media e bassa risoluzione.

La scelta di utilizzare una maschera deriva dalla considerazione che la rifocalizzazione forzata su tutti i punti dei piani wavelet causerebbe una perdita di naturalezza e profondità da parte dell'immagine.

Ricordiamo che la prima armonica wavelet è accordata sulla struttura geometrica dei bordi curvilinei mentre le armoniche di ordine superiore al primo sono accordate su strutture geometriche viva via più complesse come linee, angoli e d incroci.

Quindi risulta che il modulo dei piani wavelet di un determinato ordine dipende dalla visibilità della struttura cui è accordato, cioè quanto essa è distinguibile dalle altre presenti nell'immagine. A differenza di quanto accade per l'ordine 1, ad ogni variazione locale di luminanza corrispondono più oscillazioni sui piani wavelet, quindi si deve tener conto non solo dell'ampiezza e dello spessore delle varie tracce, ma anche della loro posizione reciproca, e risulta più difficile stabilire un'esatta corrispondenza tra l'operazione di sfocatura e le alterazioni subite dai piani.

L'algoritmo implementato elabora i piani wavelet superiori al primo, tramite coefficienti lineari $\beta_{n,j}$, scelti in base al piano wavelet e al livello di sfocatura,

per migliorare i risultati nel recupero di immagini sfocate.

L'amplificazione dei valori del piano viene effettuata solo sui punti appartenenti alle maschere M_n ottenute dai piani prodotto di ordine superiore PP_n , in formule:

$$PP_n[l][m] = \prod_{j=0}^2 |HF_{n,j}[l][m]| \quad (2.5)$$

$$M_n[l][m] = \begin{cases} 1 & \text{se } PP_n[l][m] > \gamma_n \\ 0 & \text{altrove} \end{cases} \quad (2.6)$$

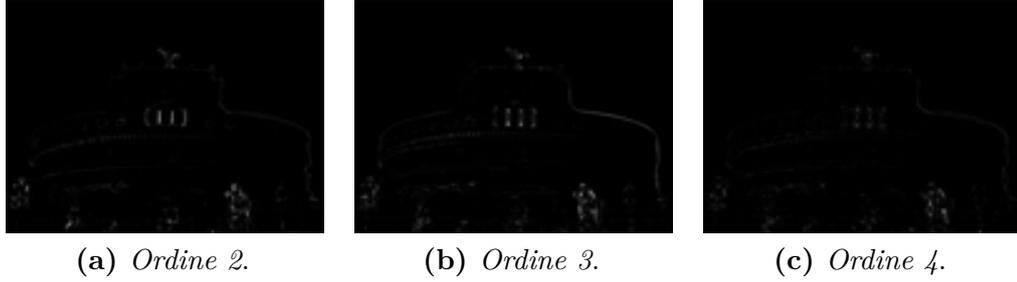


Figura 2.6. Maschere

Le maschere permettono di rifocalizzare solo le strutture principali dell'immagine, evitando, come nel caso del piano di ordine 1, artefatti indesiderati e lasciando inalterata la naturalezza dell'immagine.

Riassumendo i piani elaborati risultano

$$|HF'_{n,j}[l][m]| = \begin{cases} \beta_{n,j} \cdot |HF_{n,j}[l][m]| & \text{se } M_n[l][m] = 1 \\ |HF_{n,j}[l][m]| & \text{altrove} \end{cases} \quad (2.7)$$

con $2 \leq n \leq 4$ e $0 \leq j \leq 2$, utilizzati nel processo di ricostruzione.

Per riottenere l'immagine originale, partendo dai piani wavelet, è necessario innanzitutto ricostruire la luminanza e in seguito considerare la crominanza dell'immagine. La trasformata di Fourier della luminanza ricostruita è:

$$I = \frac{I_{LF} \cdot H_{LF}^* + \sum_{n=1}^4 \sum_{j=0}^2 (I_{n,j} \cdot H_{n,j}^*)}{|H_{LF}|^2 + \sum_{n=1}^4 \sum_{j=0}^2 |H_{n,j}|^2} \quad (2.8)$$

dove H_{LF} e $H_{n,j}$ indicano rispettivamente il filtro a bassa frequenza e i filtri ad alta frequenza di ordine n e risoluzione j . Per concludere il processo

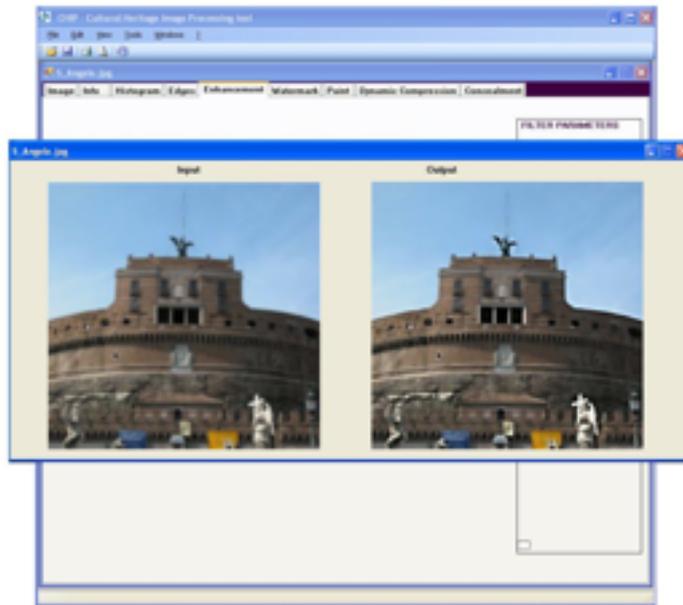


Figura 2.7. Tab Enhancement dell' applicativo CHIP

di restauro ricreiamo l'immagine a colori, riunendo la luminanza ricostruita e le crominanze U e V precedentemente estratte.

2.3 Algoritmo di compressione della dinamica

La procedura si basa sulla decomposizione a multirisoluzione mediante wavelet circolari armoniche di Laguerre-Gauss, sull'elaborazione delle sue componenti e sulla ricostruzione dell'immagine dinamicamente compressa utilizzando i piani elaborati. Immediate applicazioni delle tecniche di compressione della dinamica si hanno nella maggior parte delle applicazioni multimediali, in quanto la ridotta dinamica dei sistemi di visualizzazione oggi presenti non permettono di rappresentare immagini che presentano elevati range dinamici, con conseguente perdita della qualità di visualizzazione.

Prima di illustrare l'algoritmo sviluppato facciamo una piccola premessa sugli algoritmi basati sul contesto.

2.3.1 Tecnica omomorfica

Tale tecnica si basa sulla decomposizione di un'immagine S nelle sue componenti di riflettività R e illuminazione L , divisione fatta pixel a pixel, cosicché si può scrivere $S(x, y) = R(x, y) + L(x, y)$.

Considerando il logaritmo dell'immagine possiamo scrivere $s = l + r$ dove $s = \log(S)$, $l = \log(L)$ e $r = \log(R)$ sono i logaritmi delle singole componenti. Conoscendo i limiti del range dinamico della riflettività, poiché data a natura fisica degli oggetti essi riflettono solo parte della luce incidente, e smussamenti spaziali nel campo, possiamo imporre i vincoli $R \in [0, 1]$ e $L \geq S$, cioè $l \geq s$. Detto questo il problema che andremo a risolvere è la minimizzazione delle quantità:

$$F[l] = \int_{\Omega} (|\nabla(l)|^2 + \alpha(l - s)^2 + \beta|\nabla(l - s)|^2) dx dy \quad (2.9)$$

La 2.9 deve rispettare i vincoli $l \geq s$ e $\nabla(l) \cdot n \rightarrow 0$ su $\partial\Omega$. Dove Ω è il supporto dell'immagine mentre α e β sono numeri reali non negativi. Le condizioni necessarie e sufficienti per minimizzare $F[l]$ sono ottenute tramite le equazioni di Eulero-Lagrange:

$$\frac{\partial F[l]}{\partial l} = 0 \quad \text{per } l > s \quad (2.10)$$

da cui

$$\frac{\partial F[l]}{\partial l} = -\Delta l \alpha (l - s) - \beta \Delta (l - s) \quad l = s, \forall (x, y) \in \Omega \quad (2.11)$$

A questo punto l'algoritmo iterativo calcola i gradienti della immagini generate dalla piramide Gaussiana, partendo da un generica risoluzione p e utilizzando l'uscita dei passi precedenti come ingresso per quelli successivi, e così via analizzando tutte le possibili risoluzioni. L'algoritmo termini quando non sono più rispettati vincoli. In figura 2.8 mostriamo il risultato dell'operazione.

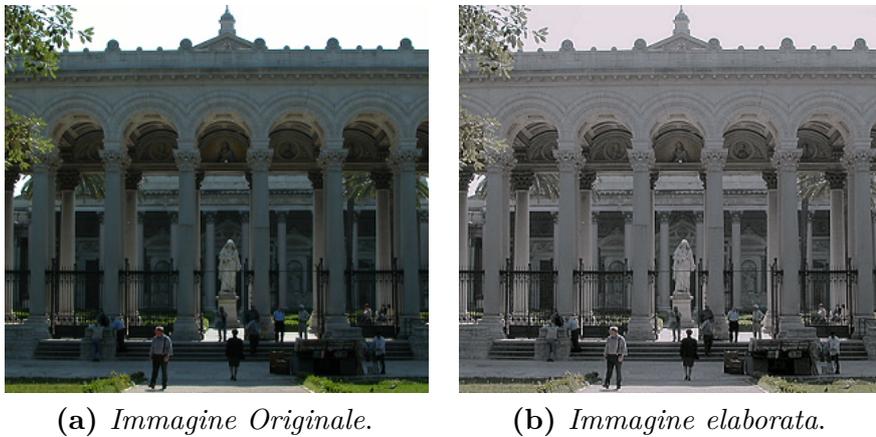


Figura 2.8. Tecnica omomorfica

2.3.2 Tecnica basata sulla compressione dei bordi

Tale metodo estrae la componente di luminanza Y dell'immagine di cui si vuole effettuare la compressione della dinamica. Definiamo la matrice come $H = \log(Y)$, ovvero il logaritmo della luminanza, quindi calcolato il gradiente di H lungo le due direzioni X ed Y , inoltre utilizziamo la decomposizione wavelet, con varianza del filtro di primo ordine, pari a zero, per le alte frequenze pari e infine consideriamo il piano wavelet di alta frequenza e massima risoluzione. A questo piano viene applicata una non-linearità del tipo:

$$\phi(x, y) = \left(\frac{\alpha}{H(x, y)} \right)^{1-b} \quad (2.12)$$

Con α pari al valor medio del piano considerato moltiplicato per un fattore $a = 0.1$ e $b = 0.9$. Il risultato così ottenuto è scomposto nelle sue componenti

reale ed immaginaria, che rappresentano rispettivamente il gradiente lungo la direzione X e Y . Adesso, ponendo $G_X = \nabla X$ e $G_Y = \nabla Y$, è calcolata, la divergenza secondo l'approssimazione delle differenze all'indietro:

$$\text{div}G \approx G_X(x, y) - G_X(x - l, y) + G_Y(x, y) - G_Y(x, y - l) \quad (2.13)$$

Viene impostato un sistema, basato sull'equazione di Poisson, in cui il Laplaciano dell'immagine è pari alla divergenza di G . Adesso il Laplaciano è approssimato sfruttando la relazione 2.14.

$$\Delta I(x, y) \approx I(x + l, y) + I(x - l, y) + I(x, y + l) + I(x, y - l) - 4I(x, y) \quad (2.14)$$

Imponendo le condizioni al contorno secondo le quali ai bordi le derivate sono nulle e utilizzando il metodo di Gauss-Seidel (appendice A.9) per la risoluzione del sistema lineare, otteniamo una matrice, che rappresenta un'immagine in cui sono maggiormente esaltate le componenti in alta frequenza. Questa immagine, insieme alla componente di bassa frequenza, estratta dall'immagine iniziale, è utilizzata per la ricostruzione wavelt. Otteniamo quindi una nuova componente di luminanza Y' nel dominio del logaritmo, a questo punto effettuando l'esponenziale di Y' otteniamo la componente di luminanza voluta Y_{new} . Infine ricostruiamo l'immagine nelle componenti R, G, B attraverso le relazioni 2.15.

$$\begin{aligned} R_{new} &= \left(\frac{R_{old}}{Y}\right) Y_{new} \\ G_{new} &= \left(\frac{G_{old}}{Y}\right) Y_{new} \\ B_{new} &= \left(\frac{B_{old}}{Y}\right) Y_{new} \end{aligned} \quad (2.15)$$

In figura 2.8 mostriamo il risultato dell'operazione.

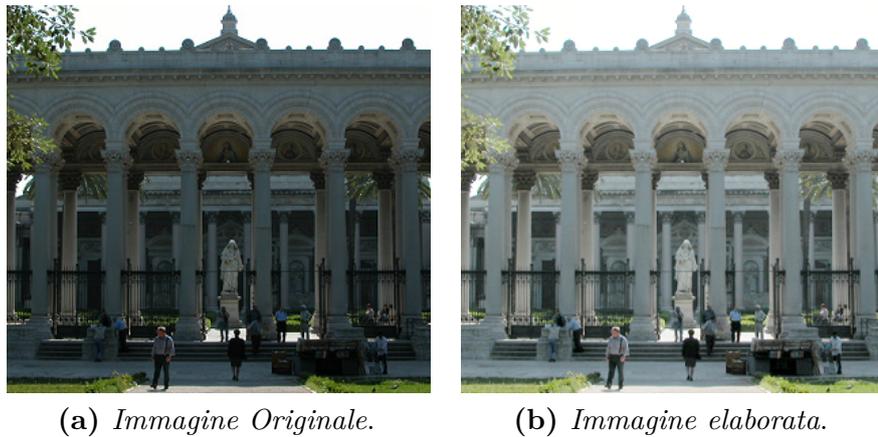


Figura 2.9. Tecnica omomorfica

2.3.3 Algoritmo High Dynamic Range

La tecnica sviluppata prende in prestito i vantaggi derivanti dalle tecniche omomorfica e compressione dei bordi, infatti da una parte utilizza la scomposizione di un'immagine nel prodotto delle sue componenti di riflettività e luminanza, dall'altra, focalizza l'attenzione sui rapidi cambiamenti del valore del gradiente dell'immagine, ottenendo un'alta attenuazione della dinamica garantendo, al tempo stesso la conservazione dei dettagli dell'immagine stessa. Il primo passo effettuato è quello di eseguire una scomposizione dell'immagine nelle sue componenti R, G, B, quindi dopo aver effettuato il passaggio dal dominio RGB a quello YUV, la componente di luminanza Y è opportunamente elaborata e utilizzata per ricostruire l'immagine. La prima operazione effettuata è il calcolo del logaritmo della luminanza $\log(1 + x)$ (figura 2.10); poiché il logaritmo della luminanza rappresenta una buona approssimazione della luminosità percepita dal sistema visivo umano.

Data un'immagine, o meglio il logaritmo della luminanza dell'immagine, consideriamo un punto generico (x, y) del piano cartesiano e definiamo kernel Gaussiano:

$$[h]g_s^0(x, y) = \left(\frac{1}{\pi s^2} \right) e^{-\frac{x^2+y^2}{s^2}} \quad (2.16)$$

dove s rappresenta un fattore di dilatazione. Definiamo un operatore gradiente complesso



Figura 2.10. Componente di luminanza $Y(a)$ e operazione di logaritmo sulla luminanza(b)

$$\begin{aligned} \nabla G_s(x, y) &= \left[\frac{d}{dx} (g_s(x, y)) + j \frac{d}{dy} (g_s(x, y)) \right] \\ &= - \left(\frac{2x}{s^2} \right) \left(\frac{1}{\pi s^2} \right) e^{-(x^2+y^2)/s^2} - j \left(\frac{2y}{s^2} \right) \left(\frac{1}{\pi s^2} \right) e^{-(x^2+y^2)/s^2}. \end{aligned} \quad (2.17)$$

L'argomento della funzione 2.17 rappresenta la direzione calcolata in ogni punto x, y . Inoltre dato l'operatore normalizzato:

$$g_s^1(x, y) = \frac{s^2}{2} \nabla G(x, y) \quad (2.18)$$

che riscritto in coordinate polari è:

$$g_s^1(r, \theta) = g_s^0(r, \theta) \left(-\frac{r}{s} \right) e^{-j\theta} \quad (2.19)$$

con $r = \sqrt{x^2 + y^2}$ e $\theta = \arctg\left(\frac{y}{x}\right)$. L'operatore 2.19 appartiene alla classe di funzione dette Funzioni Circolari Armoniche, definite come funzioni complesse, separabili sul piano polare.

Effettuando la convoluzione tra un immagine e l'operatore 2.19 otteniamo, nel dominio del tempo, la seguente relazione:

$$i_G(x, y; s) = g_s^1(x, y) * i(x, y). \quad (2.20)$$

Mentre nel dominio della frequenza

$$I_G(\xi, \eta; s) = G_s(\xi, \eta; s) * i(\xi, \eta) \quad (2.21)$$

La trasformata di Fourier $G_s(\xi, \eta)$ di $g_s^1(x, y)$ espressa in coordinate polari $\rho = \sqrt{(\xi^2 + \eta^2)}$ e $\varphi = \arctg(\eta/\xi)$ è:

$$\begin{aligned} G_s^1(\rho, \varphi) &= j2\pi \left(\frac{s}{2}\right) \rho \cdot e^{(-j\varphi)G_s^0(\rho, \varphi)} \\ &= j2\pi \cdot e^{(-j\varphi)(s/2)} e^{(\pi s \rho)^2} \end{aligned} \quad (2.22)$$

Considerata la sequenza diadica, discreta, di scala $s_k = a2^k$, con a numero reale positivo e k numero intero che indica l'ordine della risoluzione, possiamo rappresentare un'immagine utilizzando le trasformate di Fourier delle componenti alla diverse risoluzioni

$$I_{k,1}(\xi, \eta) = G_s^1(\xi, \eta) \cdot I(\xi, \eta) \quad (2.23)$$

e la componente a bassa-risoluzione

$$I_0(\xi, \eta) = G_0(\xi, \eta) \cdot I(\xi, \eta). \quad (2.24)$$

L'algoritmo utilizzando per la compressione della dinamica utilizza tre livelli di risoluzione, definiti dalla variabile k , in particolare: basso per $k = 2$, medio $k = 1$ e alto $k = 0$. La figura 2.11 mostra i piani wavelet utilizzati per la successiva elaborazione.

I piani ad alta frequenza e a diversa risoluzione, o meglio i loro moduli, sono innanzitutto normalizzati secondo la relazione:

$$\eta_k(x, y) = \frac{|i_{k,1}(x, y)|}{au_k} \quad (2.25)$$

dove u_k è la media del modulo $|i_{k,1}(x, y)|$ e a un parametro settato in fase di pre-elaborazione, e successivamente è eseguita una elevazione a potenza con esponente pari a $b - 1$, con b compreso tra 0 e 1. Notiamo che l'operazione effettuata è non lineare e fornisce in uscita i piani wavelet modificati mostrati in figura 2.12

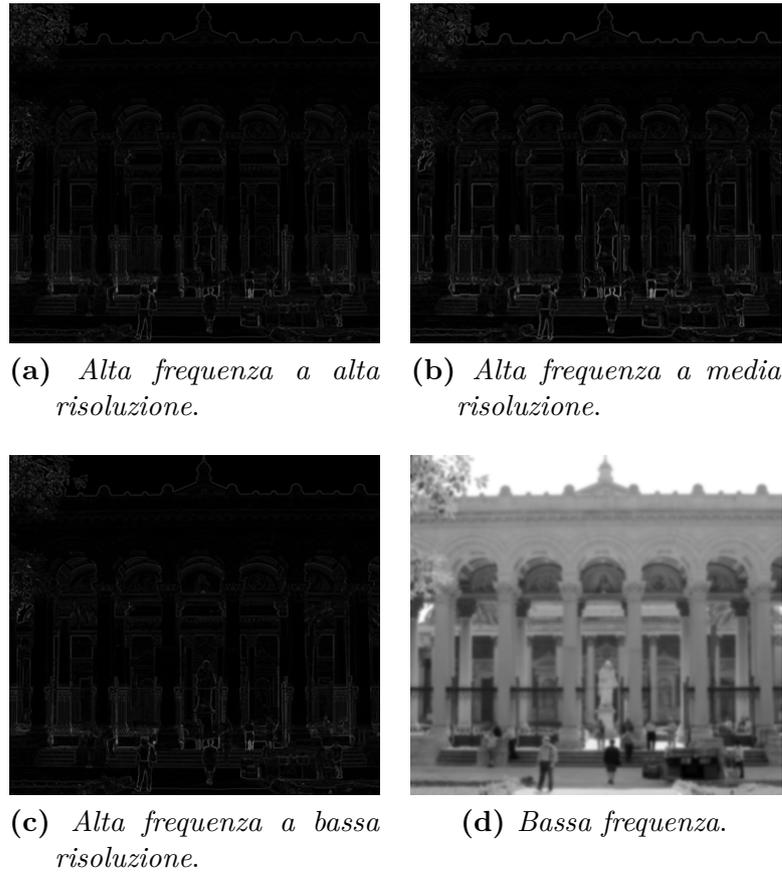


Figura 2.11. Piani Wavelet

Definiamo il fattore di attenuazione locale, mostrato in figura 2.13, del gradiente ottenuto moltiplicando pixel a pixel i piani $n_k(x, y)$, ovvero:

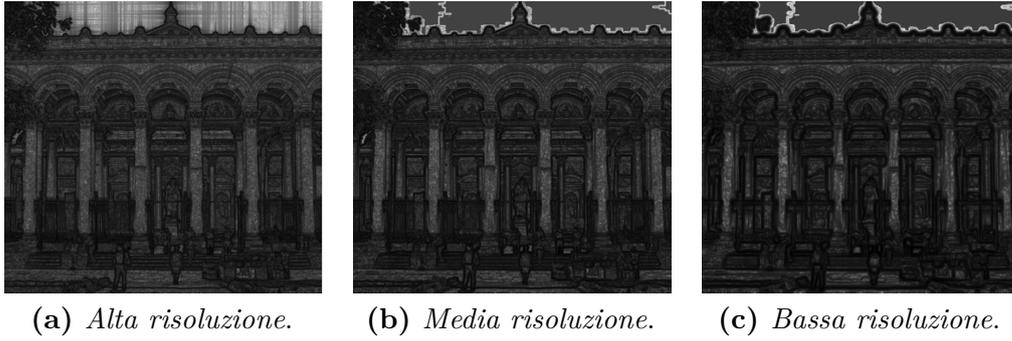
$$\psi(x, y) = [\eta_0(x, y) \cdot \eta_1(x, y) \cdot \eta_2(x, y)]^{b-1} \quad (2.26)$$

Per concludere l'elaborazione relativa le alte frequenze, moltiplichiamo il piano di alta frequenza ad alta risoluzione con il fattore di attenuazione 2.26 secondo la relazione:

$$u(x, y) = i_{0,1}(x, y) \cdot \psi(x, y) \quad (2.27)$$

Notiamo, in figura 2.14, che l'effetto ottenuto attraverso questa operazione permette di enfatizzare solo i dettagli che maggiormente influenzano la percezione visiva, senza modificare la natura stessa dell'immagine rappresentata.

Per quanto riguarda la bassa frequenza è necessario moltiplicare il piano a



(a) *Alta risoluzione.* (b) *Media risoluzione.* (c) *Bassa risoluzione.*

Figura 2.12. Non linearità applicata ai piani wavelet ad alta frequenza



Figura 2.13. Prodotto tra i piani wavelet ad alta frequenza

bassa frequenza originale per un parametro di attenuazione c ottenendone una versione modificata:

$$l(x, y) = c \cdot i_0(x, y) \quad (2.28)$$

Dopo aver ottimizzato al meglio le componenti a bassa ed alta frequenza è necessario sintetizzare l'immagine logaritmo usando una formula di ricostruzione che tenga conto delle modifiche apportate:

$$I_C(\xi, \eta) = \frac{U(\xi, \eta) \cdot G_{s_0}^{1*}(\xi, \eta) + L(\xi, \eta) \cdot G^0(\xi, \eta)}{|G_{s_0}^{1*}(\xi, \eta)|^2 + |G^0(\xi, \eta)|^2} \quad (2.29)$$

Dove $I_c(\xi, \eta)$, $U(\xi, \eta)$ e $L(\xi, \eta)$ sono rispettivamente le trasformate di Fourier dell'immagine logaritmo, del piano wavelet di alta frequenza ad alta risoluzione e della componente di bassa frequenza modificata.

La figura 2.15 mostra l'immagine logaritmo ricostruita e l'operazione espo-



Figura 2.14. Piani wavelet ad alta frequenza attenuato

nenziale eseguita su quest'ultima per l'equalizzazione dei valori.



(a) *Logaritmo ricostruito.*

(b) *Operazione esponenziale del logaritmo.*

Figura 2.15. Immagine logaritmo ricostruita dopo l'elaborazione e operazione esponenziale effettuata sull'immagine logaritmo ricostruita.

Per ottenere l'immagine a colori è necessario ricodificare le componenti R, G, B dell'immagine originale, tenendo in considerazione le elaborazioni eseguite in precedenza. Per fare questo, dette $r(x, y)$, $g(x, y)$, $b(x, y)$ le componenti R, G, B e $i(x, y)$ la luminanza dell'immagine in ingresso definiamo le nuove componenti R_C , G_C , B_C :

$$r_C(x, y) = \left[\frac{r(x, y)}{i(x, y)} \right]^s \cdot i_C(x, y) \quad (2.30)$$

$$g_C(x, y) = \left[\frac{g(x, y)}{i(x, y)} \right]^s \cdot i_C(x, y) \quad (2.31)$$

$$b_C(x, y) = \left[\frac{b(x, y)}{i(x, y)} \right]^s \cdot i_C(x, y) \quad (2.32)$$

con s parametro per la regolazione della saturazione del colore.

Unendo le componenti di colore ottenute è possibile osservare (figura 2.16, l'operazione di compressione della dinamica effettuata sull'immagine d'ingresso.



Figura 2.16. Immagine ottenuta dopo l'operazione di compressione della dinamica

L'algoritmo è stato testato su numerose immagini, ne riportiamo alcuni esempi 2.17, 2.18.

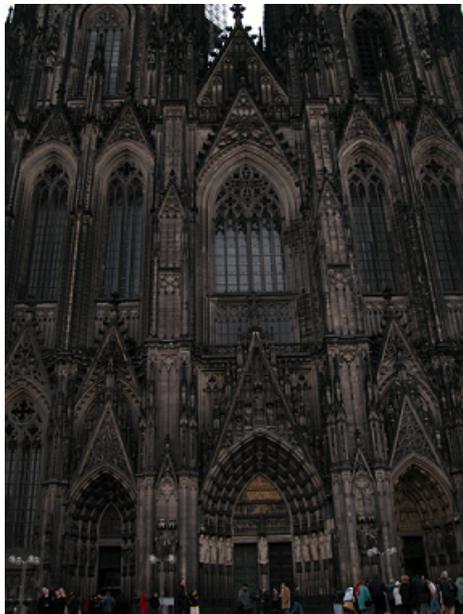


(a) *Immagine Originale.*



(b) *Immagine elaborata.*

Figura 2.17. Algoritmo HDR



(a) *Immagine Originale.*



(b) *Immagine elaborata.*

Figura 2.18. Algoritmo HDR

2.4 Digital Impainting

Nell'affrontare un metodo di restauro delle immagini si deve tenere conto delle caratteristiche tipiche del sistema visivo umano che, quasi inconsciamente, conduce ad assemblare le forme geometriche al fine di visualizzare mentalmente figure più semplici e familiari. Inoltre, considerando il completamento amodale, secondo il quale l'occhio tende ad integrare le zone mancanti di una figura, possiamo affermare che la presenza delle parti occultate si verifica senza che si sia attivato alcun modo della percezione né alcuna stimolazione prossimale. La figura 2.19 presenta tre elementi geometrici coperti, il completamento amodale ci consente di intuirne l'andamento triangolare.

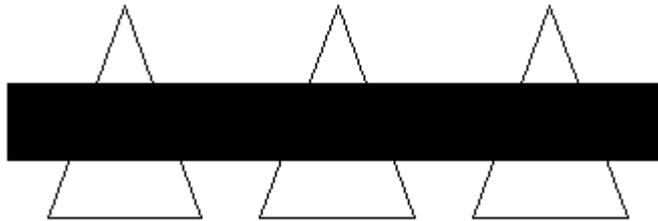


Figura 2.19. Banda nera nasconde parzialmente i triangoli

Togliendo la barra nera, scopriremo la figura 2.20 che ci fornisce i veri andamenti. Ma se si riposiziona uno schermo in corrispondenza delle anomalie la nostra esperienza non può nulla, si continua a vedere tre triangoli regolari, nonostante la nostra conoscenza a proposito della lacuna del triangolo centrale e dell'incrocio di quello a destra.

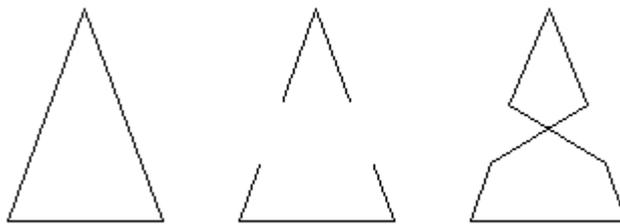


Figura 2.20. Reale forma delle figure geometriche

Un'altra interessante proprietà del completamento amodale è data dalla figura 2.21

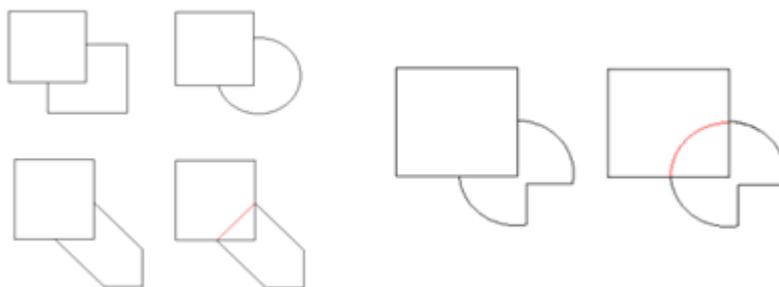


Figura 2.21. Reale forma delle figure geometriche

Qui, il completamento amodale avviene nella direzione del risultato più regolare e sembrano importanti la continuità di direzione nei punti di intersezione e la minima distanza, quando la continuità di direzione non dà luogo a figure chiuse. Una specifica tendenza alla regolarità, anche quando esistono le condizioni per il suo manifestarsi, non sembra avere un'influenza determinante. È necessario tener conto di quanto appena detto durante al fine di aiutare l'occhio a mimetizzare le zone danneggiate e a produrre una visione quanto più naturale. Dopo questa piccola premessa analizzeremo nel dettaglio l'algoritmo di Digital Impainting sviluppato per generare una simulazione virtuale del restauro di immagini relative ai beni culturali.

In particolare l'algoritmo, attraverso una accurata analisi, permette di ricostruire zone danneggiate dell'immagine recuperandone l'aspetto originale.

L'algoritmo di ricostruzione elabora principalmente la componente di luminanza dell'immagine, nel caso di immagini a colori verranno estratte dall'immagine le tre componenti RGB e la componente di luminanza Y. Dopo aver interpolato i bordi della luminanza questi ultimi saranno utilizzati dall'algoritmo di integrazione per ricostruire le tre componenti RGB.

Il procedimento avverrà in step successivi che possono essere sintetizzati in sette passi principali:

- Individuazione della zona danneggiata.
- Pre-elaborazione della zona danneggiata.
- Estrazione dei bordi.
- Inseguimento dei bordi.
- Stima delle traiettorie ed interpolazione del piano sfinato.

- Ricostruzione del gradiente dell'immagine.
- Integrazione ed eventuale ricomposizione del colore.

La prima operazione da effettuare è il rilevamento della zona da ricostruire. Esistono algoritmi automatici per il rilevamento semi-automatico dei graffi ma la scelta migliore è quella di lasciare al restauratore la selezione che potrà per esempio avvenire tramite tavoletta grafica su di un normale programma di elaborazione di immagini. Tale selezione verrà chiamata **maschera** e la zona danneggiata **lago** (figura 2.22).



Figura 2.22. Immagine danneggiata e immagine della lago

Per elaborare correttamente le informazioni contenute nei bordi è necessario eliminare la netta discontinuità in prossimità del lago, cioè la presenza di un netto contrasto che produce un bordo artificiale che non è presente nell'immagine originale non danneggiata.

Per fare ciò eseguiamo una pre-elaborazione dell'immagine al fine di riempire i primi pixel del lago con valori coerenti con quelli presenti sulla riva, utilizzando il modello dell'equazione del calore. Il modello generale matematico dell'equazione del calore 2.33 si presta bene per la ricostruzione della zona danneggiata dell'immagine, ciò è stato possibile grazie ad una modellizzazione della struttura tramite opportune equazione di diffusione.

$$D \cdot \nabla^2 I + \frac{1}{c \cdot \delta} W(x, y) = \frac{\partial I}{\partial \tau} \quad (2.33)$$

In regime stazionario e senza generazione di calore interno:

$$\nabla^2 I = 0 \quad (2.34)$$

Tale modello è utilizzato nella pre-elaborazione delle immagini al fine di eliminare i problemi di falsi bordi. L'immagine ottenuta dopo la pre-elaborazione 2.23 rappresenta la base da cui partiamo per ottenere un'immagine restaurata priva di difetti, in particolare per immagini poco contrastate questa fase permette di ottenere già la completa ricostruzione.



Figura 2.23. Immagine pre-elaborata

Eseguiamo, sfruttando l'analisi wavelet di Laguerre-Gauss, il filtraggio wavelet dell'immagine pre-elaborata e spostiamo l'elaborazione nel dominio della frequenza.

Estraiamo tre piani wavelet a differente risoluzione, low resolution (LR), medium resolution (MR) e high resolution (HR), e sottoponiamo ognuno di questi ad un processo di estrazione dei bordi sfinati, basata sullo "zero-crossing del laplaciano" dell'immagine.

Dopo tale operazione per ogni risoluzione si ottiene il relativo "piano sfinato", costituito da bordi di spessore pari ad un pixel, tale piano sarà poi messo in AND con il logo per ottenere, nel caso del piano ad HR qui utilizzato in questo lavoro, il risultato mostrato in figura 2.24.

In questi il bordo è in ogni punto univocamente definito in modulo e fase, ed è quindi possibile operare su di esso le operazioni geometriche di interpolazione. Individuiamo adesso i punti di contatto tra bordi e lacune (figura 2.25), a partire da ognuno di questi pixel si effettua quindi l'inseguimento del relativo bordo incidente al fine di estrarne le informazioni di modulo e fase che lo caratterizzano; in base ad esse si procederà poi all'interpolazione del proseguimento del bordo.



Figura 2.24. Operazione AND tra piano sfinato (HR) e lago



Figura 2.25. Immagine dei punti di confine tra bordi e lacune

La modalità di inseguimento prevede, dopo aver focalizzato il pixel da inseguire, la ricerca tra i suoi otto vicini, ad esclusione di quelli precedentemente inseguiti, quello di intensità prossima a quella del pixel considerato, cioè non superiore ad una differenza di intensità pari al 30% (figura 2.26). Per ogni bordo inseguito viene così creata una lista di punti che verrà poi immediatamente utilizzata dall'algoritmo di interpolazione.

L'algoritmo prevede che per ogni punto di incidenza trovato vengano effettuate prima la procedura di inseguimento e subito dopo quella di interpolazione della traiettoria.

L'interpolazione delle traiettoria è stata effettuata da un punto di vista geometrico, quindi è stata focalizzata l'attenzione sul modulo del piano sfinato scelto, ricostruito il quale verrà stimata la componente in fase per garantirne la



Figura 2.26. Immagine di inseguimento dei bordi del piano wavelet HR

coerenza. A tal fine per ogni sequenza di pixel inseguiti viene creato un vettore contenenti i valori delle variazioni di fase di questi ultimi, cioè individuato il pixel i -esimo P_i viene calcolato il valore $\Delta\varphi_i = \Delta\varphi_{i+1} - \Delta\varphi_i$.

I valori di questo vettore sono pressoché costanti per strutture regolari come segmenti o sezioni circolari, al contrario, per strutture aventi bruschi cambi di direzione come le configurazioni ad L, il vettore presenterà delle forti discontinuità. La lettura del vettore consente di stabilire se ricostruire il bordo per regolarità, cioè via interpolazione circolare o per direzione, cioè continuando con un segmento la traiettoria entrante nella lacuna.

Le traiettorie interpolate devono essere uniformate (figura 2.27), in termini di modulo e fase, coerentemente ai valori dei pixel inseguiti, ovvero è necessario formulare una legge che regoli l'intensità del modulo interpolato e ricostruisca le componenti in fase.

Per ricostruire il modulo consideriamo i primi 3 pixel inseguiti e ne calcolo l'intensità media, in seguito imponiamo alla relativa traiettoria ricostruita una intensità che decade con una semplice legge lineare ?? al fine di controllare al meglio la profondità di ricostruzione ed una eventuale deriva nella traiettoria stimata che condurrebbe ad un grossolano errore di ricostruzione.

$$i_x = I_0 - \frac{I_0}{N}i, \quad (2.35)$$

dove I_0 rappresenta il valore medio dei primi tre pixel prossimi alla lacuna, mentre N è il numero dei pixel interpolati, geometricamente identifica la lunghezza del bordo ricostruito o meglio il numero di pixel che sono stati scritti dentro la parte danneggiata.



Figura 2.27. Piano sfinato interpolato e piano sfinato uniformato

Utilizzando il piano sfinato interpolato, dobbiamo generare un gradiente sintetico, che sia il più possibile prossimo a quello vero, da utilizzare nel processo di integrazione. Per prima cosa filtriamo il piano sfinato ricostruito attraverso un filtro Gaussiano opportunamente amplificato, in particolare abbiamo usato:

$$G(x, y) = 3.5 \cdot \frac{\sqrt{\pi}}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (2.36)$$

Il risultato ottenuto è mostrato in figura 2.28.



Figura 2.28. Piano sfinato (HR) dopo l'applicazione del filtro

L'ultima fase dell'algoritmo prevede l'integrazione dei piani elaborati, la ricostruzione dell'immagine e la ricomposizione dei colori.

Il metodo numerico utilizzato per l'algoritmo di integrazione di Frankot-Chellappa, tale algoritmo elabora la FFT del gradiente ricostruito tramite la seguente formula 2.37.

$$\hat{I}(\omega_x, \omega_y) = \frac{a_x^*(\omega_x) \cdot \hat{G}_x(\omega_x, \omega_y) + a_y^*(\omega_y) \cdot \hat{G}_y(\omega_x, \omega_y)}{|a_x(\omega_x)|^2 + |a_y(\omega_y)|^2} \quad (2.37)$$

Dove $\hat{G}_x(\omega_x, \omega_y)$ e $\hat{G}_y(\omega_x, \omega_y)$ sono i coefficienti di un gradiente $G(\omega_x, \omega_y)$, mentre a_x e a_y sono i coefficienti del filtro di Fourier dell'operatore gradiente assunto come quello alle differenze centrali e sono pari alle relazioni 2.38.

$$\begin{aligned} a_x(\omega_x) &= \sqrt{-1} \sin(\omega_x) \\ a_y(\omega_y) &= \sqrt{-1} \sin(\omega_y) \end{aligned} \quad \text{con} \quad \begin{cases} \omega_x = \frac{2\pi}{n}u \\ \omega_y = \frac{2\pi}{n}v \end{cases} \quad (2.38)$$

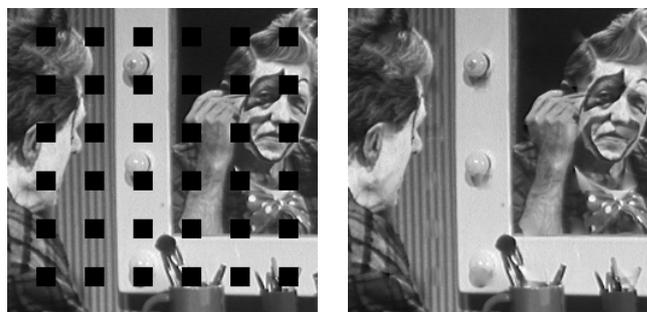
Tale elaborazione calcola i coefficienti di Fourier dell'immagine, quindi per ottenere l'immagine integrata è necessario eseguire una antitraformazione, utilizzando la formula 2.39, e infine utilizzare le componenti R, G, B, per generare l'immagine finale ricostruita e restaurata (figura).

$$I(x, y) = \sum_{u=0}^{n-1} \sum_{v=0}^{m-1} \hat{I}(u, v) \cdot e^{j\frac{2\pi(u x + v y)}{n}} \quad (2.39)$$



Figura 2.29. Ricostruzione dell'immagine

Mostriamo altre immagini su cui è stato testato l'algoritmo.

(a) *Danneggiata.*(b) *Ricostruita.***Figura 2.30.** Elaborazione immagine Farfalla(a) *Danneggiata.*(b) *Ricostruita.***Figura 2.31.** Elaborazione immagine Pepper(a) *Danneggiata.*(b) *Ricostruita.***Figura 2.32.** Elaborazione immagine Clown

2.5 Watermarking

L'utilizzo delle tecniche di digitalizzazione offre un apporto significativo alle procedure di catalogazione dei reperti storico artistici, offrendo vantaggi in termini di archiviazione e rapido reperimento delle informazioni per motivi di ricerca ad esempio. La creazione di database digitali multimediali, che contengano informazioni quali la collocazione, le caratteristiche, la datazione e lo stato di manutenzione di un bene, diventano un grande strumento per gli studiosi del settore che si trovano facilitati una serie di attività che tempo prima avrebbero richiesto settimane se non mesi di lavoro, come ad esempio i confronti incrociati. Una problematica derivante dalla grande diffusione riguarda il campo relativo alla protezione dei diritti di proprietà intellettuale e dei diritti di sfruttamento economico tutelati dalla legge. Infatti se la digitalizzazione delle immagini fornisce da un lato grossi vantaggi in termini di flessibilità nella fruizione e nel reperimento delle informazioni, dall'altro presenta problemi di tutela degli archivi digitalizzati da attacchi di pirateria sempre in agguato quando si parla di accesso telematico

Queste motivazioni hanno spinto il mondo della ricerca a sviluppare algoritmi che operino in tale direzione. In particolare il digital watermarking consente di applicare un identificativo univoco ad una immagine o ad un contenuto multimediale introducendo all'interno degli stessi dati informazioni relative all'autore dell'opera, al suo proprietario, al distributore e addirittura al suo fruitore finale allo scopo di tutelare e garantire l'utilizzo del contenuto in un ambito di perfetta legalità e rispetto dei diritti di proprietà e di distribuzione. Le procedure di watermarking vengono applicate per una ampia gamma di prodotti e rispondono ad esigenze di varia natura dalla garanzia di autenticità di un prodotto o servizio, ai diritti di riproduzione e di utilizzo. Per comprendere la struttura del nostro lavoro, pertanto, pare opportuno in questa sede fare una breve panoramica dei metodi di classificazione delle procedure di watermarking. Si suole classificare il watermarking in base a tre parametri:

- **Tecnica di inserimento** del marchio: è detto visibile se il marchio è visivamente percettibile (ad esempio filigrana delle banconote, o i canali televisivi) o invisibile se il contenuto informativo inserito è impercettibile e la qualità del bene multimediale subisce degradazioni molto limitate.

- **rilevamento del marchio:** è definito blind se per identificare il watermark all'interno del documento marchiato non occorre né l'immagine originale né il marchio, no-blind se nella fase di rilevazione sono necessari sia l'immagine originale sia il marchio, infine semi-blind se la rilevazione è effettuata utilizzando solo il marchio.
- **dominio di applicazione:** in questo caso si ha marchiatura spaziale se l'operazione è effettuata direttamente sui pixel che costituiscono l'immagine, o marchiatura nel dominio della frequenza se l'operazione di marchiatura è realizzata dopo che l'immagine ha subito una trasformazione spazio-frequenza, come la trasformata coseno (DCT), trasformata di Fourier o trasformata wavelet.

Quando parliamo di marchiatura digitale, ovvero di digital watermarking dobbiamo tenere in considerazione che la sua maggiore efficacia nella protezione dei diritti di proprietà intellettuale si robusto, impercettibile invertibile e avere una bassa probabilità di errore, in altri termini deve essere univocamente determinabile.

La caratteristica di robustezza indica la capacità della marchiatura digitale di resistere agli attacchi che hanno lo scopo di modificarlo, sostituirlo o eliminarlo in maniera più o meno intenzionale, questo in altri termini significa che l'identificativo dovrebbe codificare una tale quantità di informazioni da renderlo praticamente inattaccabile, sarebbe auspicabile, inoltre, che una marchiatura elettronica non fosse facilmente individuabile, in modo da non deteriorare la qualità del documento marchiato e al tempo stesso da non renderne agevoli e semplificate le alterazioni. D'altro canto non è attualmente pensabile creare un "marchio forte", ovvero inattaccabile in termini di protezione dei diritti di proprietà senza incidere, con un ragionevole margine di probabilità, in negativo sulla integrità, in termini ad esempio di qualità dell'immagine, del documento stesso.

La scelta delle caratteristiche del marchio, del metodo di inserimento e della porzione di documento in cui esso deve essere inserito, rappresentano un punto nodale durante le fasi di progettazione di un qualsiasi sistema di marchiatura e sono strettamente connesse al tipo di prodotto cui la marchiatura è diretta.

2.5.1 Tecniche di watermarking

Un processo di watermarking, a seconda del campo di applicazione, deve soddisfare differenti requisiti e, di conseguenza, deve implementare determinate tecniche di inserimento ed estrazione del marchio. In particolare per applicazioni relative alla protezione del copyright, la bontà di un sistema di watermarking viene misurata considerando i seguenti parametri:

- **Invisibilità** della modifica, allo scopo di non diminuire il valore commerciale dell'opera.
- **Robustezza** del marchio nei confronti di possibili attacchi da parte di utenti maliziosi.
- **Complessità** computazionale efficiente.

Ogni processo di watermarking può essere schematizzato ad alto livello come composto da due sotto-processi fortemente correlati tra loro : encoding e decoding, o anche detti di inserimento e verifica. L'encoding può essere rappresentato con uno schema a blocchi 2.33.



Figura 2.33. Schema a blocchi del codificatore

Data una immagine digitale X , detta cover, e dato il watermark $W = \{w_1, w_2, w_3, \dots, w_N\}$ costituito da una sequenza binaria, l'immagine marchiata risultante I_W è ottenuta utilizzando una funzione di codifica :

$$C(X, W) = X_W \quad (2.40)$$

dove il marchio potrebbe dipendere dalle caratteristiche dell'immagine. Per completezza riportiamo due tipiche funzioni utilizzate per l'inserimento del marchio 2.41.

$$\begin{aligned}x'_i &= x_i + \alpha \cdot w_i \\x'_i &= x_i (1 + \alpha \cdot w_i)\end{aligned}\tag{2.41}$$

dove x'_i, x_i e w_i rappresentano, rispettivamente, l' i -esimo valore dell'immagine marchiata, originale e del marchio. Il parametro di scala α indica la 'forza' del marchio, cioè quanto esso incide sulla qualità dell'immagine, permette quindi di adattare il processo di marchiatura digitale alle caratteristiche dell'immagine.

Nella fase di decoding (figura 2.34) è necessaria una funzione di decodifica D che permetta di astrarre dall'immagine marchiata l'informazione relativa al marchio. In particolare, nel caso di metodo privato o non cieco la funzione è data dalla relazione $D(X_w, X) = W'$

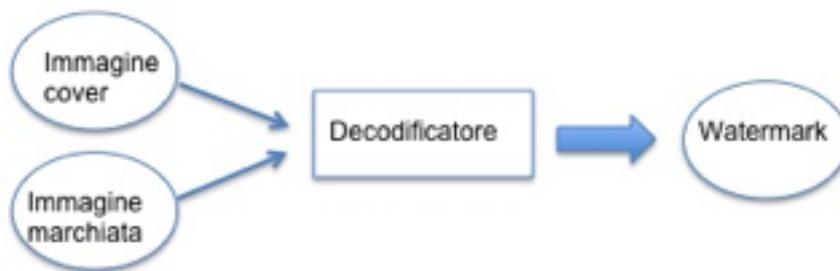


Figura 2.34. Schema a blocchi del decodificatore no-blind

Mentre, nel caso di metodo pubblico o cieco (figura 2.35) la relazione è $D(X_w) = W'$



Figura 2.35. Schema a blocchi del decodificatore blind

Come si può notare, la differenza risiede nella conoscenza, da parte del decoder, dell'immagine originale (cover). Il watermark, estratto dopo la decodifica W' , viene comparato con il watermark originale, inserito durante la fase di codifica, utilizzando una funzione di comparazione che valuti la similarità tra i due marchi.

2.5.2 Dominio del watermarking

I sistemi di watermarking permettono di definire due classi principali relative al dominio in cui operano i processi di inserimento e rilevazione del marchio. Parleremo di dominio spaziale di un'immagine quando si agisce direttamente sui valori dai pixel che la costituiscono. In questo caso le operazioni di gestione del watermarking risultano immediate e facilmente realizzabili, ma è altresì vero che le caratteristiche di visibilità e robustezza del marchio raggiungono livelli inaccettabili.

Citiamo come esempio di watermarking spaziale un algoritmo, semplice e comunemente utilizzato, denominato Less Significant Bits (LSB), che permette di modificare i bit meno significativi di ogni pixel dell'immagine.

Consideriamo l'immagine "Lena", rappresentata in scala di grigio tramite una matrice di pixel, dove ognuno di essi è una stringa di bit che indica l'intensità di grigio.

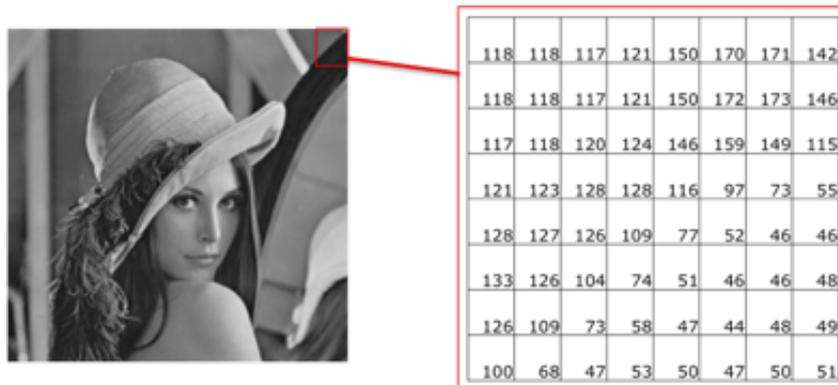


Figura 2.36. Watermarking nel dominio spaziale

Dato il marchio da inserire nell'immagine w_1, \dots, w_n , sostituiamo l'ultimo bit dell' i -esimo pixel p_i con l' i -esimo valore del marchio:

$$p_i = 110101 \rightarrow p'_i = 11010w_i \quad (2.42)$$

L'altro metodo utilizzato per l'inserimento del watermark prevede l'utilizzo di tecniche che agiscono nel dominio delle frequenze, e si basa sull'applicazione di filtri lineari alla trasformata dell'immagine considerata. Quelle comunemente utilizzate sono: la Trasformata Coseno Discreta (DCT)(vedi appendice A.7) e la Trasformata Wavelet Discreta (DWT)(vedi appendice A.8).

Tutte le trasformate utilizzate sono invertibili, pertanto dopo l'applicazione

della trasformazione diretta e inversa è possibile riottenere esattamente la stessa informazione iniziale.

In generale considerata una immagine digitale, di dimensione $N \times M$, una trasformazione lineare invertibile può essere definita dalla relazione 2.43.

$$F(u, v) = \sum_{x=1}^N \sum_{y=1}^M f(x, y) \cdot A(x, y, u, v) \quad (2.43)$$

dove $f(x, y)$ rappresenta il valore del pixel, in posizione $f(x, y)$, prima della trasformazione, e $A(x, y, u, v)$ rappresenta la base dell'operatore di trasformazione invertibile.

La trasformazione inversa restituisce il valore del pixel originario, in posizione $f(x, y)$, utilizzando i valori trasformati $F(u, v)$, la cui equazione è data dalla 2.44.

$$f(x, y) = \sum_{u=1}^N \sum_{v=1}^M F(u, v) \cdot B(x, y, u, v) \quad (2.44)$$

con $u = 1, \dots, N$, $v = 1, \dots, M$ e $B(x, y, u, v)$ base dell'operatore di trasformazione inverso. Il procedimento di marchiatura può essere riassunto in tre passi

1. calcolo della trasformata dell'immagine,
2. inserimento del watermark in opportuni coefficienti della trasformata
3. calcolo della trasformata inversa per tornare al dominio spaziale.

Prestiamo attenzione che risulta fondamentale la scelta dei coefficienti della trasformata che conterranno il marchio, poiché questo determina la bontà marchiatura digitale, essa dipende dalla trasformata utilizzata e dalle caratteristiche intrinseche dell'immagine. In particolare si può dimostrare che l'energia media dei coefficienti DCT risulta assai spesso sbilanciata, cioè viene data maggiore rilevanza ai primi coefficienti, che rappresentano le basse frequenze, rispetto agli ultimi, che rappresentano le alte frequenze.

Sfruttando sia la proprietà di compattazione dell'energia della DCT, sia le caratteristiche del sistema visivo umano, che risulta maggiormente sensibile alle variazioni delle basse frequenze, con contenuto visivo rilevante, rispetto alle alte frequenze, è possibile realizzare un efficiente algoritmo per la marchiatura digitale delle immagini. In pratica considerata una immagine e la sua

trasformata DCT (figura 2.37), notiamo che i coefficienti posizionati in alto a sinistra rappresentano la bassa frequenza, mentre quelli in basso a destra rappresentano le alte frequenze.

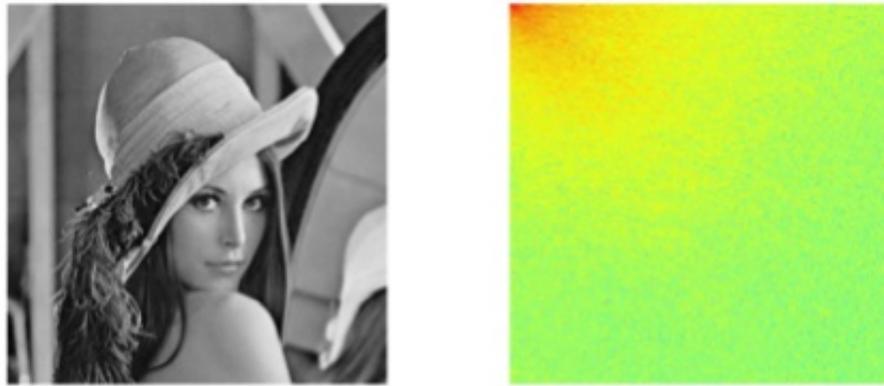


Figura 2.37. Immagine Lena e la corrispondente trasformata DCT

Per realizzare un algoritmo di watermarking robusto ed invisibile è necessario scansionare la matrice dei coefficienti DCT partendo dal primo valore in alto a sinistra fino ad arrivare all'ultimo in basso a destra secondo un percorso a 'zig-zag', ed introdurre il valori del nostro marchio nei coefficienti DCT che rappresento le medie frequenze. Per quanto riguarda la trasformata wavelet, descritta in appendice A.8, sembra maggiormente predisposta per la marchiatura digitale. Osservando le figure 2.39 che mostrano, rispettivamente, l'immagine originale e la trasformata wavelet, possiamo fare alcune considerazioni a riguardo.

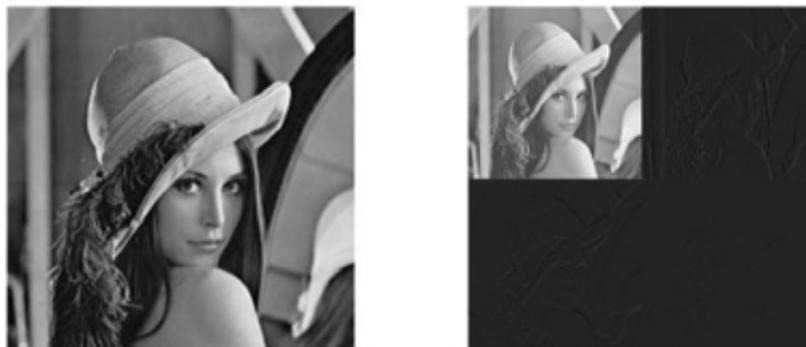


Figura 2.38. Immagine Lena e la corrispondente trasformata DCT

Le bande LH, HL e HH (figura ??) conterranno quindi le alte frequenze presenti nell'immagine, che essendo meno percettibili al sistema visivo umano sono ottime candidate per l'inserimento del marchio digitale. Le sottobande risultano resistere, in maniera differente, ai diversi tipi di attacco cui è sottoposta l'immagine marchiata.

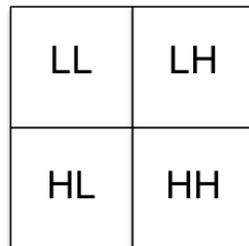


Figura 2.39. Suddivisione in bande di frequenza della trasformata wavelet.

È necessario considerare questo aspetto nella fase di inserimento del marchio per migliorare la robustezza e l'efficienza del nostro algoritmo. In particolare l'inserimento del marchio nella banda LL, basse frequenze, è più resistente a compressione lossy, rumore Gaussiano, ridimensionamento, rotazione, blurring, ritaglio e sharpening. Nelle bande HL, LH e HH, invece, l'inserimento del marchio è resistente istogramma equalization, alla correzione della gamma e dell'intensità. Infine l'inserimento del marchio in tutte le sottobande permette di affrontare attacchi di re-watermarking e collusione. Ovviamente bisogna considerare a priori, l'aspetto qualitativo dell'immagine, infatti la presenza del marchio in zone ad alto contenuto informativo dell'immagine, come la banda LL, causa un peggioramento della qualità dell'immagine stessa. L'algoritmo di marchiatura digitale sviluppato, utilizzando le wavelet circolari armoniche (vedi appendice A.1), inserisce il marchio scegliendo in maniera opportuna i coefficienti della trasformata, appartenenti alle diverse sottobande di frequenza, al fine di mantenere più alta possibile la qualità stessa dell'immagine. L'algoritmo di tipo no-blind, quindi conosciamo il marchio, effettua una marchiatura digitale delle immagini molto robusta e permette di recuperare il marchio stesso anche in situazioni di manomissione dell'immagine stessa a causa di errori dovuti alla trasmissione o di attacchi mirati.

Preso in considerazione l'immagine Lena (figura 2.40), abbiamo testato la bontà del nostro algoritmo valutando la qualità delle immagini marchiate attraverso l'uso dell'indice di qualità oggettiva Peak Signal-to-Noise Ratio (PSNR) tabella

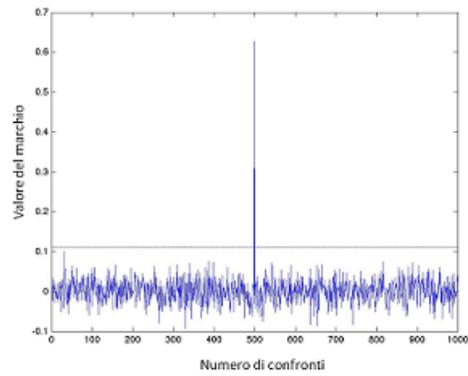
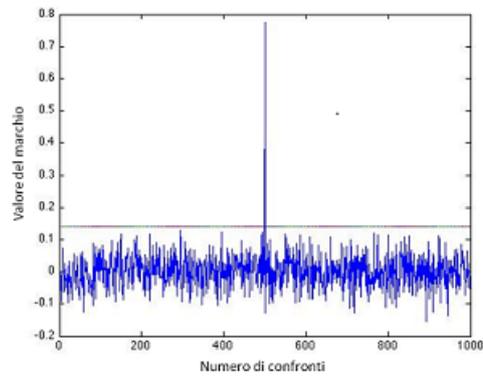
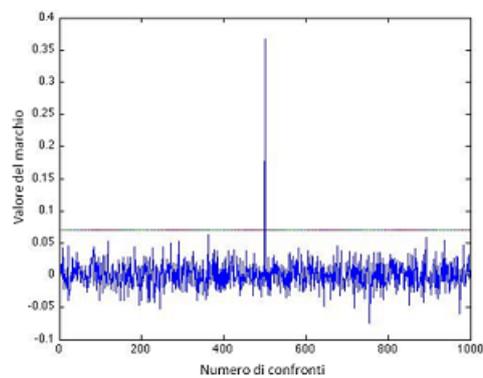
2.1). Inoltre nelle figure 2.412.422.432.442.452.46 sono mostrate le immagini corrotte dagli attacchi e i relativi grafici che indicano la rilevazione del marchio tra un numero considerevole di possibili candidati.

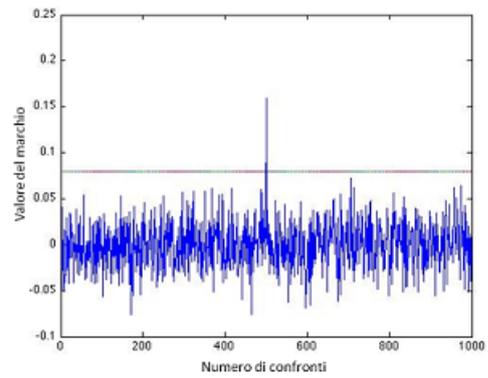
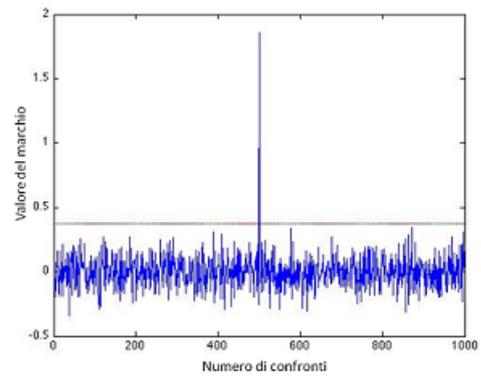
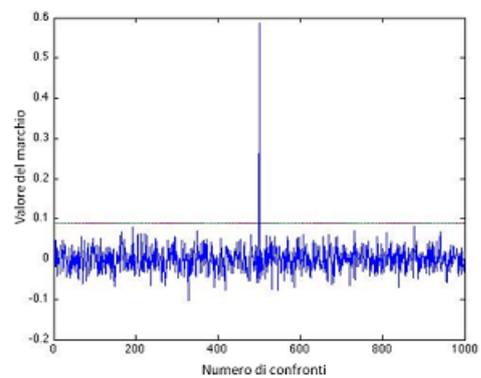


Figura 2.40. Immagine Lena.

Tipo di attacco	PSNR
Solo marchiatura	51.79
Equalizzazione istogramma	18.78
Filtraggio Gaussiano	33.34
Compressione JPEG (40%)	34.70
Sharpening	25.60
Rumore	21.35

Tabella 2.1. Tabella riassuntiva degli attacchi e relativi valori di PSNR

(a) *Marchiata.*(b) *Rilevatore del marchio.***Figura 2.41.** Immagine contenete il marchio(a) *Attacco.*(b) *Rilevatore del marchio.***Figura 2.42.** Equalizzazione dell'istogramma(a) *Attacco.*(b) *Rilevatore del marchio.***Figura 2.43.** Filtraggio Gaussiano

(a) *Attacco.*(b) *Rilevatore del marchio.***Figura 2.44.** Compressione JPEG (40%)(a) *Attacco.*(b) *Rilevatore del marchio.***Figura 2.45.** Sharpening(a) *Attacco.*(b) *Rilevatore del marchio.***Figura 2.46.** Rumore additivo

Capitolo 3

Structure From Motion

3.1 Introduzione

Per “digitalizzazione 3D” intendiamo il processo di cattura della geometria di un oggetto reale mediante l’uso di macchinari appositi. In letteratura troviamo moltissimi termini che rappresentano questo stesso concetto quali: fotogrammetria digitale, 3d scanning, range acquisition, shape acquisition, 3d model acquisition, rangefinding e range scanning.

Le sue applicazioni sono molteplici e aumentano di anno in anno man mano che la tecnologia permette realizzazioni sempre più veloci, precise e automatiche. Queste spaziano dall’ambito medico a quello della sicurezza, dalla conservazione del patrimonio artistico all’intrattenimento visuale.

La possibilità di trasformare i dati in un modello 3d, permette all’operatore di analizzare un qualsiasi oggetto come se lo avesse in mano. Attualmente, esistono diversi metodi di digitalizzazione 3d, principalmente essi si caratterizzano in base al fatto che l’acquisizione necessita o meno del contatto fisico con la superficie in esame (figura 3.1).

I **metodi attivi** sfruttano la proprietà della luce che viene proiettata in modo strutturato (pattern luminosi, luce laser, radiazioni infrarosse, ecc) e, misurando la distorsione che essa subisce una volta riflessa si determina la forma dell’oggetto. Il modo in cui viene calcolata l’informazione di profondità determina la sottoclasse di appartenenza del digitalizzatore. Elenchiamo ora alcune delle tecniche attive utilizzabili: active defocus, stereo attivo, triangolazione attiva, interferometria, tempo di volo (ToF). I **metodi passivi** non utilizzano sorgenti ausiliarie bensì la luce naturale riflessa dall’oggetto.

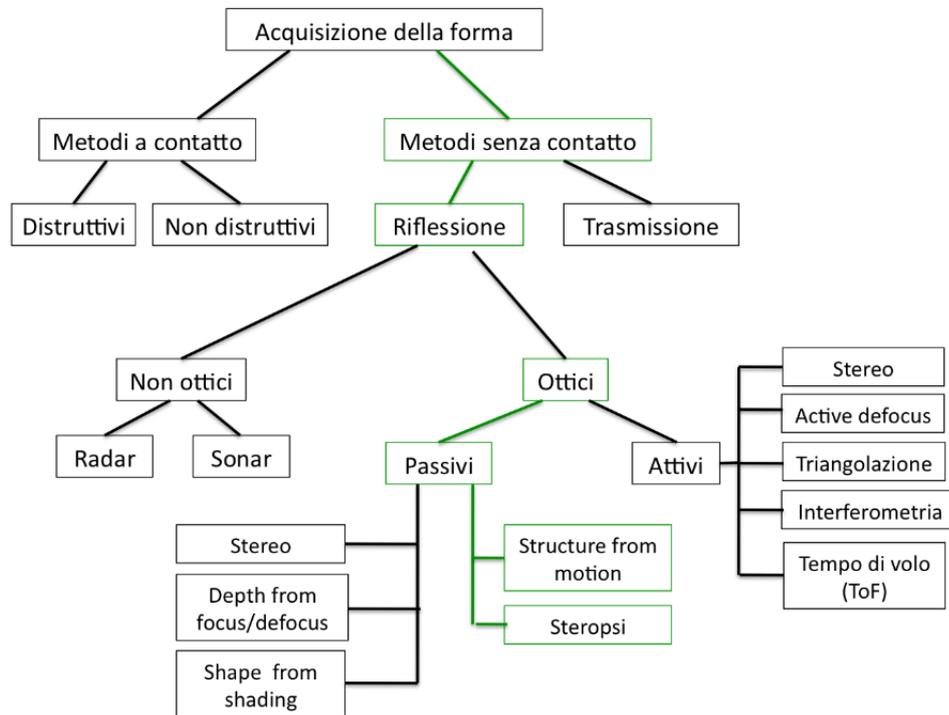


Figura 3.1. Classificazione acquisizione metodi ottici

Solitamente le caratteristiche di questa sorgente sono sconosciute. Da notare quindi che queste tecniche non interagiscono in modo attivo con l'oggetto in esame né tramite un contatto fisico, né tramite un irraggiamento. L'analisi dell'oggetto può essere quindi effettuata in base ad informazioni prelevate anche da persone non addette ai lavori, che hanno semplicemente fotografato l'oggetto in questione, infatti è sufficiente che siano presenti tante più foto possibile scattate da diversi punti di vista. Questa caratteristica fa sì che possano essere ricostruiti oggetti di cui si possiede solo un'informazione fotografica. Inoltre, l'economicità dell'hardware e la velocità nel prelevare i dati contraddistinguono questi metodi da tutti gli altri. Ovviamente questi pregi vanno a discapito della precisione del modello e tempo di calcolo. Gli algoritmi preposti a tale compito sono molto complessi e possono impiegare molte ore di elaborazione per un singolo oggetto. La precisione invece è strettamente legata alla risoluzione di acquisizione dell'immagine, ed allo stato attuale non raggiunge i livelli di alcuni sensori ottici attivi.

Alcune tecniche passive utilizzate sono: depth from focus/defocus, shape from texture, shape from shading, stereo fotometrico, stereopsi, shape from silhouette,

shape from photo-consistency, structure from motion. Prima di descrivere alcune tecniche, maggiormente usate nell'ambito dei metodi ottici passivi, ovvero la Stereoscopia, la Stereopsi, la Structure From Motion, analizziamo il funzionamento del sistema visivo umano e di un apparato per l'acquisizione di immagini, per comprendere come sia possibile invertire tale processo ed ottenere così la geometria tridimensionale di una scena da semplici immagini digitali.

3.1.1 Fisiologia del sistema visivo umano

I nostri occhi sono composti da un bulbo e da alcuni annessi come i muscoli estrinseci, che servono per effettuare la rotazione del bulbo stesso all'interno della sua sede, le ghiandole, le vie lacrimali e le palpebre, che servono per la lubrificazione della sclera. Il bulbo, di forma grossolanamente sferica, è costituito da tre pareti chiamate tonache che, procedendo dall'esterno verso l'interno, sono la sclera, la coroide e la retina; quest'ultima ospita i recettori della luce (figura 3.2).

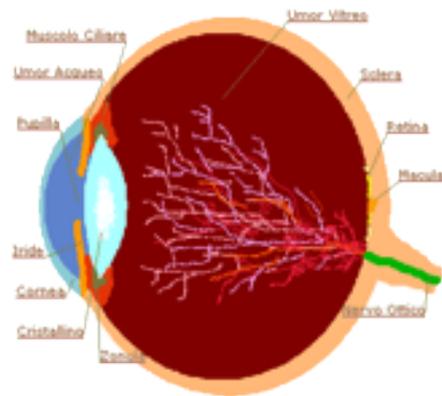


Figura 3.2. Sezione dell'occhio

L'immagine è focalizzata e impressa sulla retina dal sistema di "lenti" composto dai mezzi diottrici dall'esterno verso l'interno: la cornea, l'umor acqueo, il cristallino ed il corpo vitreo, di questi solo il cristallino ha la possibilità di deformarsi, su sollecitazione dei muscoli intrinseci, per ottenere l'esatta focalizzazione dell'immagine sulla retina. L'iride è un diaframma che possiede nella sua parte centrale la pupilla elemento dell'organo con calibro variabile.

In questo modo l'iride regola, in seguito ad un riflesso detto fotomotore, l'intensità di luce che raggiunge la retina entro limiti ottimali, come ad esempio nell'occlusore di una macchina fotografica.

I fotorecettori della retina, che si differenziano in coni e bastoncelli, raccolgono lo stimolo luminoso e lo trasformano in eccitazione nervosa tramite reazioni biochimiche. Questo stimolo viene poi convogliato, tramite il nervo ottico, ai centri corticali, deputati alla visione, presenti nel cervello.

Nonostante l'uomo possieda due occhi, non vede doppio grazie al processo di fusione, che permette il riconoscimento degli oggetti osservati ed è quindi fondamentale per l'apprendimento e, conseguentemente, per la conoscenza. L'immagine di ogni punto visibile costituente un oggetto osservato, chiamato punto oggetto, viene proiettata sulla retina in una coppia di punti, uno per ciascun occhio, chiamati punti retinici. Quando si fissa su un oggetto, gli assi ottici oculari convergono, intersecandosi in un particolare punto chiamato di fissazione. La sua immagine proiettata sulle retine, in particolare sulle fovee, si forma su una coppia di punti retinici, uno per ciascun occhio, dai quali originano stimoli nervosi che il sistema della visione "fonde" in un singolo punto.

Per questo motivo tali punti retinici sono chiamati corrispondenti; non sono simmetrici da un punto di vista anatomico, ma risultano funzionalmente accoppiati a causa del fenomeno della fusione sensoriale. Per ogni punto di fissazione esiste una curva, chiamata oroptero, dal greco *oros*, limite e *opter*, osservatore, composta da tutti i punti dello spazio reale per i quali si verifica la fusione. Tutti i punti dello spazio reale giacenti sull'oroptero vengono percepiti come punti singoli.

Per riconoscere gli oggetti osservati è necessario che questi siano percepiti come unici, la corteccia celebrale continuamente governa la direzione di osservazione tramite i movimenti degli occhi, affinché le immagini cadano su punti retinici corrispondenti. La fusione non avviene solamente per i punti giacenti sull'oroptero, ma anche per quelli che si trovano in una ristretta zona, chiamata area di Panum (figura 3.3).

I centri preposti alla visione, analizzando le disparità retiniche tra le sue immagini sulle retine, e quelle relative al punto di fissazione. I punti oggetto che si trovano al di fuori dell'area di Panum vengono visti come doppi, fenomeno chiamato diplopia. Le immagini di tali punti oggetto si formano su punti

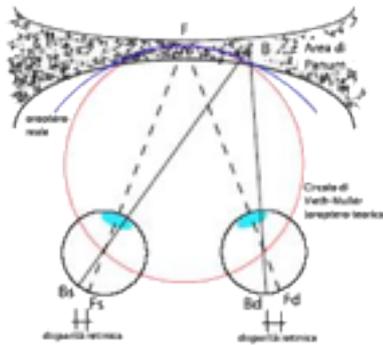


Figura 3.3. Schema della visione

retinici che, non essendo accoppiati, si chiamano disparati e non portano alla fusione.

Per renderci conto di questo fenomeno possiamo effettuare un semplice esperimento, mettiamo davanti ai nostri occhi ad una certa distanza l'indice della mano sinistra. Se lo fissiamo lo vedremo singolo; se poniamo dito l'indice della mano destra tra il nostro volto e il punto che stiamo fissando, vedremo quest'ultimo come doppio (diplopia crociata) ed identica sensazione avremo se lo posizioneremo tra l'indice della mano sinistra e l'infinito (diplopia omonima). Se un oggetto percepito come doppio è posto tra gli occhi e l'oroptero parleremo di disparità crociata mentre se è posto tra l'oroptero e l'infinito parleremo di disparità omonima. Abituamente non ci accorgiamo della diplopia cosiddetta fisiologica poiché la nostra attenzione è focalizzata sull'oggetto che stiamo osservando e quindi il nostro cervello trascura le immagini doppie.

Dopo questa breve introduzione sul funzionamento del sistema visivo umano introduciamo il concetto di stereopsi, ovvero la visione tridimensionale originata dalla stimolazione simultanea di punti retinici orizzontalmente disparati da parte di punti oggetto situati nell'ambito delle aree di Panum. Definiamo inoltre acutezza stereoscopica la disparità minima oltre la quale non si produce alcun effetto stereoscopico. La percezione della profondità, non dipende quindi, solamente dalla visione binoculare, ma è frutto di un insieme di elaborazioni mentali e di fenomeni percettivi e fisiologici la cui conoscenza permette di ricreare artificialmente la sensazione di distanza.

Le arti visive, in particolare la pittura con la prospettiva, la fotografia e la cinematografia, hanno spesso sfruttato tali processi, a volte anche effettuando artificiose elaborazioni sul materiale visivo, per ottenere la sensazione di di-

stanza desiderata. Esistono elementi monoculari che concorrono al giudizio della distanza relativa tra gli oggetti osservati, ad esempio:

- Il movimento parallattico: la velocità di spostamento di un oggetto vicino sembra maggiore di quella di uno lontano.
- La prospettiva lineare: un oggetto di grandezza costante sottende angoli progressivamente minori man mano che si allontana, quindi è visto più piccolo. L'esempio classico è dato dalla convergenza apparente di due rette parallele che si allontanano dall'osservatore.
- La sovrapposizione dei contorni: un oggetto che interrompa i contorni di un altro viene percepito come anteposto all'altro.
- La distribuzione delle luci e delle ombre: il chiaro-scuro genera una impressione di rilievo e quindi di profondità.
- La familiarità con oggetti noti: la distanza tra oggetti noti viene giudicata anche in base alla loro grandezza apparente.
- La prospettiva aerea: la foschia atmosferica influenza il contrasto e il colore degli oggetti situati più lontano. Inoltre ne sfuma i contorni. Il fenomeno viene spesso utilizzato in fotografia mettendo a fuoco il soggetto principale e sfocando ad arte gli altri piani, in particolare lo sfondo.

Gli elementi binoculari invece aggiungono precisione e qualità alla percezione della distanza. La stereopsi sfrutta le piccole differenze tra le immagini recepite dai due occhi, ovvero le disparità retiniche la cui analisi e misura da parte del sistema percettivo conferiscono all'oggetto impressione di solidità e di profondità.

Adesso analizziamo il funzionamento di un apparato per l'acquisizione di immagini, per comprendere come sia possibile invertire tale processo ed ottenere così la geometria tridimensionale di una scena da semplici immagini digitali.

3.1.2 Modelli geometrici della fotocamera

Iniziamo descrivendo la camera a foro stenopeico costituita da una scatola a tenuta di luce dotata su un lato di un piccolo foro (foro stenopeico), dal quale possono entrare i raggi luminosi riflessi dal soggetto che vanno a formare

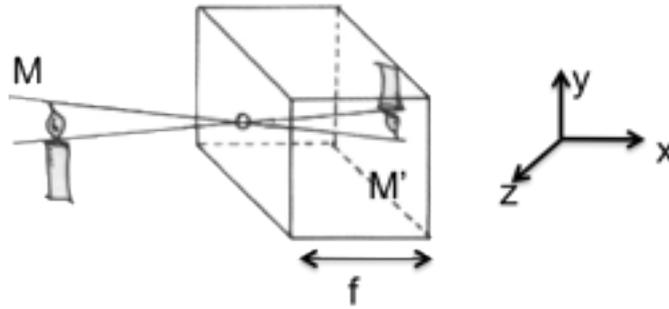


Figura 3.4. Schema camera pinhole

un'immagine invertita sul lato opposto. Dentro la camera, sul lato opposto al foro, viene collocato un foglio di materiale fotosensibile al buio. Dopo avere inquadrato il soggetto, il tappo, utilizzato per tappare il buco, viene aperto per un determinato tempo di esposizione e i raggi di luce che attraversano il foro permettono di “impressionare il foglio di materiale fotosensibile, posto sul lato opposto, che registra l'immagine “latentedel soggetto (figura 3.4).

Considerato il punto M di coordinate (X, Y, Z) e definito il punto M' , di coordinate (X', Y', Z') , come la sua proiezione sul piano immagine, e detta distanza focale f del foro dal piano immagine, otteniamo, considerando la similitudine dei rettangoli le relazioni 3.1.

$$\frac{X'}{f} = \frac{X}{Z} \quad e \quad \frac{Y'}{f} = \frac{Y}{Z} \quad (3.1)$$

da cui otteniamo le relazioni 3.4.

$$X' = \frac{-f \cdot X}{Z} \quad Y' = \frac{-f \cdot Y}{Z} \quad Z' = -f \quad (3.2)$$

Si dimostra quindi che l'immagine è invertita rispetto alla scena, come indicato dal segno negativo, inoltre la presenza di Z al denominatore implica che la dimensione dell'immagine di un oggetto è inversamente proporzionale dalla distanza di questo dall'osservatore.

Adesso complichiamo il modello sostituendo il piccolo foro con una lente convergente, che per semplicità approssimiamo ad una lente sottile, otteniamo il seguente schema dove i raggi paralleli (figura 3.5),

dove l'immagine di un punto della scena, distante Z dalla lente, viene “messo a fuoco ad una distanza dalla lente che dipende dalla profondità Z del punto e dalla distanza focale D della lente.

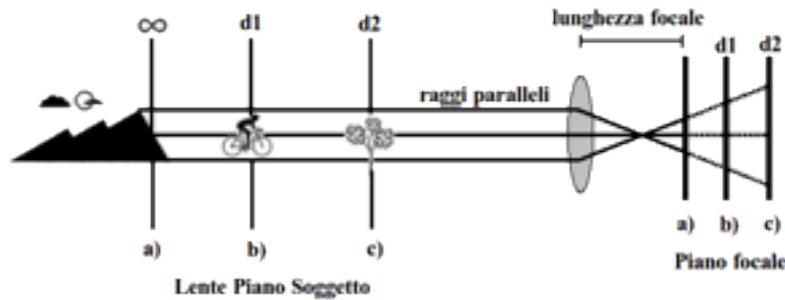


Figura 3.5. Rapporto tra distanza focale e distanza soggetto

Notiamo che tutti i punti di un soggetto posti alla stessa distanza dalla lente formano il piano bidimensionale soggetto, quindi nello spazio si possono avere infiniti piani soggetto, paralleli tra loro e perpendicolari all'asse ottico della lente, che si avvicinano e si allontanano dalla lente. Ogni piano soggetto viene focalizzato dalla lente solo sul relativo piano focale posto dietro alla lente ad una distanza che è dettata dall'equazione della lente sottile.

Riportato quanto detto nella vita reale consideriamo una fotocamera digitale che risulta composta dall'ottica, approssimabile con una coppia di lenti sottile e una matrice CCD o CMOS che costituisce il piano immagine.

Il piano immagine è definito come una matrice $n \times m$ di celle fotosensibile, che convertono l'intensità della radiazione luminosa che vi incide in impulso elettrico, mentre la matrice CCD o CMOS è convertita in un'immagine digitale, ovvero una matrice di valori interi a cui corrisponde un determinato colore.

Chiameremo gli elementi della matrice pixel e indicheremo con $I(u, v)$ il valore dell'immagine nel pixel individuato dalla riga v e colonna u .

Poiché la dimensione della matrice CCD non necessariamente coincide con quella della matrice di pixel dobbiamo considerare la relazione ?? che tiene in considerazione le dimensioni efficaci del pixel.

$$u_{pix} = \frac{N}{n} u_{CCD} \quad v_{pix} = \frac{M}{m} v_{CCD} \quad (3.3)$$

3.1.3 Stereosopia

Stereoscopia significa “visione spaziale”, e deriva dalle parole greche “stereo” che significa “spazio” e “skopein” che vuol dire “vedere. Negli ultimi decenni sono state inventate decine di tecniche ognuna delle quali è sicuramente valida in una determinata situazione. Tuttavia oggi non esiste una tecnica chiaramente superiore alle altre, soprattutto se si vuol far coincidere la qualità della visione stereoscopica e l’esigenza di spendere poco. Sfruttare le varie proprietà della luce visibile, cioè luminosità, frequenza d’onda e polarizzazione e le capacità dell’apparato visivo umano, cioè visione stereoscopica e i limiti fisici dell’occhio, rappresentano l’unica parte della ricerca sulla stereoscopia artificiale sulla quale è possibile lavorare per rendere l’esperienza più realistica possibile.

Possiamo classificare le tecniche stereoscopiche nelle classi:

- Stereoscopia attiva.
- Stereoscopia passiva.
- Autostereoscopia.

Con il termine **stereoscopia attiva** si considerano tutte le tecniche stereoscopiche che fanno ricorso all’utilizzo di occhialini 3D contenenti parti elettroniche che intervengono nel processo di visione stereoscopica. Questa famiglia di tecniche stereoscopiche è caratterizzata da un alto grado nella qualità delle immagini ottenute, tuttavia l’uso di questi occhialini è costoso. Un fattore importante da considerare in questa tecnologia è che spesso le immagini proiettate sullo schermo sono alternate, un frame è dedicato all’occhio sinistro e nell’istante successivo un frame per l’altro occhio, quindi è necessario un ulteriore dispositivo hardware che metta in sincronia la trasmissione delle immagini stereoscopiche e gli occhialini 3D, questo per evitare che l’immagine captata dall’occhio destro sia quella proiettata per l’occhio sinistro e viceversa. A questa classe appartengono diverse tecnologie, per completezza le elenchiamo riportando una breve descrizione.

La tecnologia shutter: utilizza due piccoli schermi LCD montati al posto delle lenti il segnale di controllo è generato da un circuito posizionato nella montatura degli occhiali che riceve il segnale di sincronizzazione proveniente da un trasmettitore direttamente collegato al computer o al proiettore che si occupa di generare la proiezione stereoscopica (figura 3.6).



Figura 3.6. Occhiali con tecnologia shutter

Il primo modo di alternare le immagini è detto page flipping, in questo caso il frame per l'occhio destro è pronto e visualizzato sullo schermo, il pannello LCD degli occhialini oscurerà l'occhio sinistro e viceversa in accordo con il segnale di sincronizzazione. Questa tecnica garantisce il massimo della qualità video in quanto ogni immagine ha la stessa qualità e risoluzione, tuttavia il punto dolente della tecnica è che richiede di essere visualizzata tramite un elevato refresh.

Il secondo modo è detto interlacing, in questo caso le immagini sono proiettate tramite le linee pari e poi quelle dispari alternamente, gli occhialini shutter si occuperanno poi di far visualizzare correttamente i frame ai corrispettivi occhi. La qualità è più che sufficiente per soddisfare l'occhio umano ed è praticamente obbligatoria da utilizzare se si vogliono ricreare visioni stereoscopiche nell'ambito di trasmissioni televisive.

Il terzo modo detto sync doubling divide lo schermo in due parti, la metà superiore e la metà inferiore, all'interno di una di queste parti viene visualizzata una delle immagini della coppia di immagini stereoscopiche. La qualità delle immagini così ottenute non è elevata in quanto la risoluzione verticale rispetto ad una tecnica di page flipping è dimezzata, tuttavia il costo computazionale da eseguire sul computer è assai minore. La tecnologia dei visori 3D utilizza dei caschi, meglio noti come Head Mounted Display (HMD), dotati di un consistente apparato hardware in grado di fornire visioni stereoscopiche e sonoro ad elevatissima immersività (figura 3.7).



Figura 3.7. Head Mounted Display

Un HMD dispone di una camera oscura posizionata di fronte gli occhi dell'utente dove al suo interno sono posizionati due schermi LCD, uno per ciascun occhio, con una risoluzione abbastanza elevata e una profondità di campo orizzontale fino a 65° . É inoltre dotato di delle cuffie stereo ed un dispositivo tracker che fornisce informazioni riguardanti l'orientamento della testa dell'utente al computer che si occupa di creare la visione stereoscopica. Con il termine **stereoscopia passiva** tendiamo ad indicare tutta quella famiglia di tecnologie che utilizzano occhiali 3D senza alcuna dotazione di hardware complesso, atti a fornire visioni stereoscopiche. In questa famiglia di tecniche la visione stereoscopica è ricreata utilizzando dei filtri polarizzati o particolari proprietà della luce visibile per veicolare ogni immagine della coppia di immagini stereoscopiche verso il giusto occhio. Le immagini della coppia stereoscopica sono fornite all'apparato visivo dell'utente contemporaneamente, cioè entrambi gli occhi ricevono la corrispettiva immagine nello stesso istante, e la sincronizzazione delle immagini non è gestita da un hardware ad hoc. La qualità delle immagini ottenute tramite queste tecniche non è elevata come le immagini ottenute tramite tecniche di stereoscopia attiva, tuttavia il rapporto tra qualità, costo e soprattutto numero di utenti che si possono coinvolgere nella visione stereoscopica è molto elevato. Anche in questo caso, come per la stereoscopia attiva, esistono diverse tecnologie. L'anaglifa: definisce un frame attraverso degli oggetti dette anaglifi, costituiti dalla sovrapposizione dell'immagine destra e dell'immagine sinistra della coppia di immagini stereoscopiche con due diversi colori, ad esempio il rosso per l'immagine destra e il blu per l'immagine sinistra, che idealmente rappresentano il punto di vista dei due occhi dell'osservatore della scena.

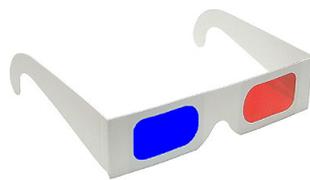


Figura 3.8. Occhiali per visualizzazione anaglifi

Osservando l'immagine tramite le lenti di colori analoghi (figura ??) si ottiene che l'occhio destro veda la sola immagine destra e l'occhio sinistro la sola immagine sinistra. Questa tecnica ha il vantaggio che entrambi i canali stereoscopici, sinistro e destro, possono venire rappresentati su un singolo

anaglifo con grande risparmio economico e semplicità realizzativa. Gli anaglifi in genere mancano di contrasto e sono completamente saturi rispetto ai colori utilizzati.

La stereoscopia cromatica: utilizza occhialini utilizzati composti da una lente formata da tanti prismi affiancati che hanno il compito di creare un effetto di diffrazione sulla luce visibile. L'immagine utilizzata è creata appositamente utilizzando palette di colori ben precisa e studiata, infatti questa tecnica si basa sulla proprietà di certi colori che se affiancati ad altri sembrano essere più vicini o lontani. Lo svantaggio che difficilmente si può ricreare artificialmente una qualsiasi scena del mondo reale con i giusti colori, però risulta efficace nella rappresentazione scientifica di dati o pubblicità, con la possibilità di fornire informazioni aggiuntive su concetti come la profondità e la distanza anche senza usare gli occhialini 3D.

La tecnica della luce polarizzata lineare sfrutta le caratteristiche delle onde elettromagnetiche e utilizza dei filtri dotati di polarizzazione. In particolare quando un fascio di luce non polarizzato passa ad esempio attraverso un filtro polarizzante orizzontale; quest'ultimo bloccherà la componente orizzontale della luce non polarizzata e lascerà passare la componente verticale della luce che ora è diventata polarizzata. Utilizzando questa tecnica è possibile visualizzare contemporaneamente le immagini stereoscopiche purché queste abbiano una polarizzazione diversa tra loro, e l'utente sia dotato di un paio di occhiali polarizzati (figura 3.9) in grado far ricevere ad ogni occhio la corrispondente immagine. Gli inconvenienti risiedono nella scarsa luminosità delle immagini a causa della presenza dei filtri polarizzati e nell'allineamento prefissato dei filtri che impedisce all'utente di muovere eccessivamente la testa altrimenti l'effetto tridimensionale diminuisce di molto fino a scomparire.

La tecnica della luce polarizzata circolare riesce ad evitare il difetto che si presenta quando l'utente china la testa a destra o sinistra e non riesce a vedere più nessuna immagine tridimensionale. Infatti utilizzando dei filtri circolari non c'è il bisogno che i filtri posti davanti ai proiettori e quelli dell'utente debbano essere allineati.

L'autostereoscopia si basa sull'utilizzo di monitor particolari che permettono di ottenere visioni stereoscopiche senza l'utilizzo di occhialini 3D, tale effetto è ottenuto con un normale monitor LCD a cui è stata aggiunta una lens plate (figura 3.10) progettata e costruita per quello specifico modello. La lens plate è



Figura 3.9. Occhiali polarizzati

una lente che riflette la luce in maniera tale che ogni occhio dell'utente possa vedere una serie di righe di pixel prefissate, si basa sul fenomeno detto barriera di parallasse, ovvero l'immagine percepita dipende dall'angolo con cui si guarda il monitor.

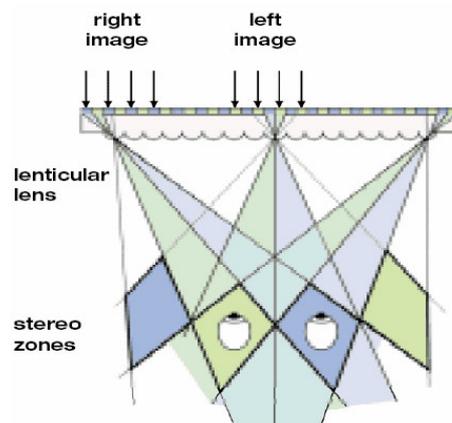


Figura 3.10. Tecnologia lens plate

In questo modo gli occhi di un osservatore vedranno due immagini diverse che opportunamente calcolate daranno vita alla visione stereoscopica. Lo svantaggio principale resta il limite posto dal Lens plate che è in grado di proiettare solo in certe direzioni le immagini stereoscopiche e quindi solo in certe posizioni frontali allo schermo l'apparato visivo sarà in grado di vedere immagini tridimensionali. Per migliorare tale tecnica si utilizza una telecamera dotata di un software in grado di riconoscere, tracciare ed inseguire la posizione degli occhi dell'utente è possibile costruire uno schermo autostereoscopico con una zona stereo mobile che segue lo spostamento della testa dell'utente. A partire da queste premesse è stato avviata una indagine approfondita e un intenso studio delle tecniche 3D e correlate problematiche, come ad esempio la

gestione della profondità di campo, della convergenza, o l'eccessiva variazione cromatica e espositiva, al fine di sviluppare algoritmi di elaborazione per la correzione di tali difetti e il miglioramento della qualità visiva.

Questi algoritmi potranno essere inseriti in un sistema dedicato alla fruizione di contenuti multimediali nell'ambito dei beni culturali. Migliorando così il grado di soddisfazione dell'utente (user experience).

3.2 La Stereopsi

3.2.1 Principi base della Stereopsi

Dopo aver affrontato il funzionamento del sistema visivo umano, il modello geometrico di visione e la tecnica della stereoscopia introduciamo la tecnica alla base della rappresentazione 3d: la Stereopsi.

La stereopsi è la visione tridimensionale che origina dalla stimolazione simultanea di punti retinici orizzontalmente disparati da parte di punti oggetto situati nell'ambito delle aree di Panum. La percezione della profondità, non dipende quindi, solamente dalla visione binoculare, ma è frutto di un insieme di elaborazioni mentali e di fenomeni percettivi e fisiologici la cui conoscenza permette di ricreare artificialmente la sensazione di distanza.

La Stereopsi è realizzata sfruttando un processo che consente di ottenere informazioni sulla struttura tridimensionale di un modello a partire da due o più immagini digitali realizzate da fotocamere che inquadrano la scena da posizioni diverse, per ovvi motivi dovuti alla natura stessa del sistema visivo umano, cioè la distanza interoculare (5-7 cm), queste non possono essere molto distanti.

Per comprendere i concetti alla base della stereopsi definiamo il modello geometrico di una fotocamera (figura ??)

dove R è il piano immagine, F è il piano focale, parallelo al piano f , C è il centro ottico ed f è la distanza focale. La retta passante per C e perpendicolare al piano R è detto asse ottico.

Definiamo due sistemi di riferimento, il primo per lo spazio tridimensionale detto anche sistema mondo, nel quale l'asse Z coincide con l'asse ottico, e il secondo detto piano di riferimento con gli assi u e v orientati come X e Y .

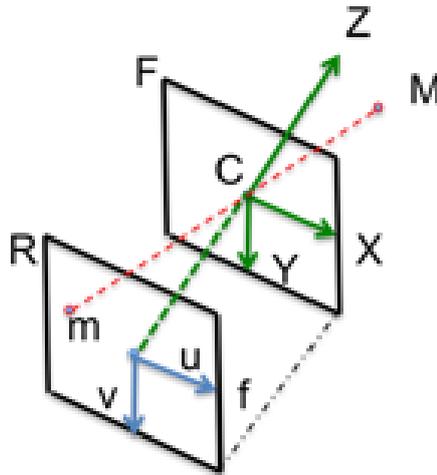


Figura 3.11. Modello generico fotocamera

Detto M , in coordinate cartesiane, un punto nello spazio tridimensionale e m la sua proiezione su R attraverso C possiamo scrivere la relazione 3.4

$$\tilde{M} = [x, y, z]^T \quad \tilde{m} = [u, v]^T \quad (3.4)$$

Data la figura 3.12 consideriamo la similitudine dei triangoli

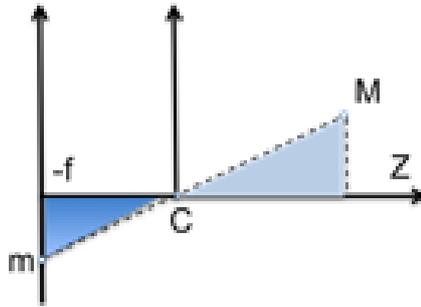


Figura 3.12. Similitudine triangoli del modello generico della fotocamera

notiamo che valgono le relazioni 3.5, che rappresentano la proiezione prospettica, ovvero le trasformazione dalle coordinate tridimensionali a quelle bidimensionali.

$$\frac{f}{z} = \frac{-u}{x} = \frac{-v}{y} \quad \left\{ \begin{array}{l} u = \frac{-f}{z}x \\ v = \frac{-f}{z}y \end{array} \right. \quad (3.5)$$

Utilizzando le coordinate omogenee 3.6 otteniamo l'equazione prospettica 3.7.

$$m = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad M = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (3.6)$$

$$z = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} -fx \\ -fy \\ z \end{bmatrix} = \begin{bmatrix} -f & 0 & 0 & 0 \\ 0 & -f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (3.7)$$

L'equazione 3.7 scritta in notazione matriciale a meno di un fattore di scala sarà $m \cong PM$.

P è la matrice di proiezione prospettica (MPP), che può essere riscritta in forma compatta 3.8.

$$P = \begin{bmatrix} p_1^T \\ p_2^T \\ p_3^T \end{bmatrix} \quad (3.8)$$

da cui otteniamo le relazioni 3.9.

$$P = \begin{bmatrix} p_1^T M \\ p_2^T M \\ p_3^T M \end{bmatrix} = \begin{bmatrix} p_1^T \\ p_2^T \\ p_3^T \end{bmatrix} M \quad e \quad \begin{cases} u = \frac{p_1^T M}{p_3^T M} \\ v = \frac{p_2^T M}{p_3^T M} \end{cases} \quad (3.9)$$

Inoltre definiamo parametri intrinseci sono i parametri necessari a collegare le coordinate di un pixel dell'immagine con le coordinate corrispondenti nel sistema di riferimento della camera.

Quindi considerando una trasformazione affine che tiene conto della traslazione del centro ottico e la riscalatura indipendente degli assi u e v

$$\begin{cases} u = k_u \frac{-f}{z} x + u_0 \\ v = k_v \frac{-f}{z} y + v_0 \end{cases} \quad (3.10)$$

Otteniamo quindi la relazione 3.11.

$$P = \begin{bmatrix} -fk_u & 0 & u_0 & 0 \\ 0 & -fk_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = K [I/0] \quad (3.11)$$

dove K è 3.12.

$$K = \begin{bmatrix} -fk_u & 0 & u_0 \\ 0 & -fk_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.12)$$

I parametri **intrinseci** sono $-fk_u, -fk_v$ e rappresentano la lunghezza focale in pixel orizzontali e verticali rispettivamente, u_0, v_0 .

I parametri **estrinseci** sono i parametri che definiscono la posizione ed orientazione del sistema di riferimento della camera, rispetto al riferimento mondo, che è supposto noto. Per legare i due sistemi di riferimento è necessario introdurre un cambio di coordinate composto da una rotazione R ed una traslazione t . valgono le relazioni 3.13.

$$M_S = GM \quad \text{con} \quad G = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \quad (3.13)$$

La matrice G (equazione 3.13) contiene i parametri estrinseci della fotocamera, ovvero tre relativi alla rotazione e tre alla traslazione rispettivamente. Considerata la rappresentazione generale della matrice prospettica 3.8 otteniamo la relazione 3.14, in cui compaiono i parametri intrinseci K , i parametri estrinseci G e la matrice di proiezione prospettica.

$$P = K[I/0]G \quad (3.14)$$

Un ultimo aspetto da affrontare è il concetto vincolo epipolare, il quale afferma che un punto in una immagine corrisponde ad una retta nello spazio e tale retta coincide sulla seconda immagine ad una retta detta epipolare. Quindi il punto omologo non va cercato sull'intera immagine ma solo sulla retta epipolare 3.11.

Questa considerazione permette di trasferire la ricerca delle corrispondenze al caso unidimensionale invece che bidimensionale, semplificando notevolmente le cose.

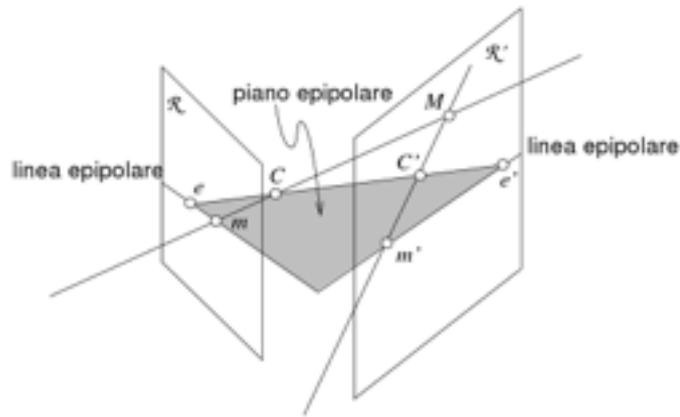


Figura 3.13. Modello tridimensionale della scena con vincoli epipolari

Dalla figura notiamo che il punto m' è vincolato a giacere sul piano epipolare generato da m . La retta CC' prende il nome di baseline.

Fissato un sistema di riferimento assoluto e date le due matrici di proiezione prospettica $P = [Q/q]$ e $P' = [Q'/q']$, otteniamo le relazioni

$$\begin{cases} m \cong PM \\ m' \cong P'M \end{cases} \quad (3.15)$$

La retta epipolare (figura 3.14) corrispondente ad m è la proiezione secondo P' del raggio di m , che ha equazione 3.16.



Figura 3.14. Linee epipolari

$$M = C + \lambda \begin{bmatrix} Q^{-1}m \\ 0 \end{bmatrix} \quad (3.16)$$

poiché $P'C = P' \begin{bmatrix} Q^{-1}q \\ 1 \end{bmatrix} = q' - Q'Q^{-1}q = e'$ e $P' \begin{bmatrix} Q^{-1}q \\ 1 \end{bmatrix} = Q'Q^{-1}m$, la retta epipolare di m ha equazione 3.17, e rappresenta l'equazione, in coordinate omogenee, della retta passante per i punti e' e la proiezione del punto all'infinito $Q'Q^{-1}m$.

$$m' \cong \lambda Q'Q^{-1}m + e' \quad (3.17)$$

Moltiplicando entrambi i membri dell'equazione 3.17 per la matrice antisimmetrica $[e']_X$, che rappresenta il prodotto esterno di e' otteniamo la relazione

$$[e']_X m' \cong \lambda [e']_X Q'Q^{-1}m \quad (3.18)$$

Moltiplicando l'equazione 3.18 per m'^T e poiché la parte a sinistra è un vettore ortogonale a m' otteniamo la relazione 3.19, che è detta equazione di Longuet-Higging e rappresenta una forma bilieare di m e m' .

$$0 = m'^T \lambda [e']_X Q'Q^{-1}m \quad (3.19)$$

Infine definiamo **matrice fondamentale** $F = [e']_X Q'Q^{-1}$, che contiene tutta l'informazione relativa alla geometria polare.

Possiamo quindi riscrivere l'equazione di Longuet-Higging come $m'^T F m = 0$

3.2.2 Calcolo della Stereopsi

La Stereopsi a livello applicativo è realizzata mediante un processo che consente di ottenere informazioni sulla struttura tridimensionale di un modello a partire da due immagini digitali realizzate da fotocamere che inquadrano la scena da posizioni diverse, ovviamente la posizione e l'angolo di ripresa tra due immagini non possono differire molto.

La tecnica di stereopsi è costituita da due fasi principali:

- Ricerca delle corrispondenze.
- Triangolazione.

Il calcolo delle corrispondenze consiste nel trovare coppie di punti coniugati in due immagini, ovvero che sono proiezione dello stesso punto della scena.

Definiamo disparità tra due punti coniugati è la loro differenza se le due immagini fossero sovrapposte e assumiamo che le due immagini non siano troppo diverse, quindi vi siano molte parti simili nelle due immagini. Infine, prima di procedere, dobbiamo introdurre il processo di rettificazione delle MPP che trasforma le MPP delle due fotocamere in modo da rendere i piani focali coplanari e i punti coniugati abbiano le stesse coordinate verticali 3.15.

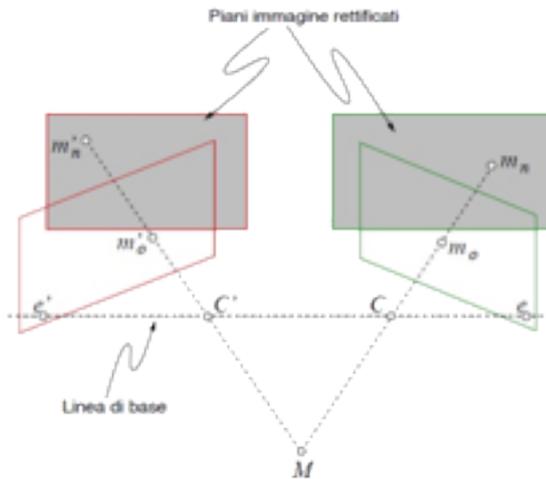


Figura 3.15. Schema rettificazione

Dopo la rettificazione è come se avessimo immagini scattate da due fotocamere con stessi parametri intrinseci e orientamento, l'unica differenza è la posizione dei centri ottici che rimangono gli stessi delle vecchie camere ??.



Figura 3.16. Schema rettificazione

La problematiche a cui andiamo incontro sono legate alla natura stessa delle immagini:

- **False corrispondenze:** un punto dell'immagine può essere messo in corrispondenza a molti punti dell'altra immagine.
- **Occlusioni:** per un punto dell'immagine può non esistere il coniugato nell'altra.
- **Distorsione radiometrica:** intensità percepite dalle due fotocamere diverse a causa di superfici non perfettamente Lambertiane.
- **Distorsione prospettica:** a causa della proiezione prospettica, gli oggetti hanno forme diverse.

Per nostra fortuna ci vengono incontro alcuni vincoli molto importanti:

- **Somiglianza:** alcuni punti appaiono simili nelle due immagini.
- **Geometria epipolare:** il punto coniugato giace su una retta (epipolare).
- **Liscezza:** lontano dai bordi la variazione in profondità varia lentamente.
- **Unicità:** un punto dell'immagine ha al più un punto coniugato.
- **Ordinamento monotono:** considerati due punti coniugati m_1 e m'_1 , allora se m_2 giace a destra (sinistra) di m_1 , allora anche il coniugato m'_2 giace a destra (sinistra) di m'_1 (Non vale se m_2 è nel cono d'ombra di m_1)

Imposti tali vincoli differenziamo i metodi di accoppiamento in globali, che estendono i vincoli ad un numero elevato di pixel, e locali, che impongono il vincolo ad un intorno del pixel da accoppiare.

I metodi locali, che presentano un minor costo computazionale, e consentono una ricostruzione 3D densa. In pratica astraggono dall'immagine immagine piccoli blocchi rettangolari e cercano blocchi simili nell'altra mediante misure di somiglianza. e sfruttando il vincolo epipolare è possibile restringere la ricerca per blocco alle rette epipolari. Tra criteri di somiglianza comunemente usati ricordiamo quelli **basati sulle differenze di intensità** dei pixel presenti nella regione analizzata, ovvero Sum of Squared Differences (SSD 3.20) e Sum of Absolute Differences (SAD 3.21).

$$SSD(u, v, d) = \sum_{k,j} (I_1(u+k, v+l) - I_2(u+k+d, v+l))^2 \quad (3.20)$$

$$SAD(u, v, d) = \sum_{k,j} |I_1(u+k, v+l) - I_2(u+k+d, v+l)| \quad (3.21)$$

In letteratura esistono anche criteri accoppiamento **basati sulla correlazione** (equazione 3.22).

$$NCC(u, v, d) = \frac{\sum_{(k,j)} I_1(u+k, v+l) - I_2(u+k+d, v+l)}{\sqrt{(I_1(u+k, v+l) - I_2(u+k+d, v+l))^2}} \quad (3.22)$$

I molti casi potremmo andare incontro a problemi in corrispondenza di discontinuità di profondità, che cerchiamo di eliminare adottando finestre adattive/eccentriche e metodi multirisoluzione. Inoltre in caso di occlusioni è necessario sfruttare vincolo di ordinamento e il vincolo destra-sinistra, secondo il quale se p è accoppiato con p' effettuando la ricerca da I_1 a I_2 , allora p' deve essere accoppiato a p effettuando la ricerca da I_2 a I_1 .

Oltre ai criteri basati sulla somiglianze dobbiamo citare i metodi **basati sulle features** delle immagini, che estraggono i punti salienti dalle due immagini e cercano di accoppiare tale features sfruttando i loro descrittori per valutarne la similarità e i vincoli per escludere falsi accoppiamenti. L'unico inconveniente è che non consentono ricostruzione densa, ma sparsa, quindi vanno eventualmente combinati con metodi di ricostruzione densa.

Per concludere il discorso sulla stereopsi esaminiamo la seconda fase del processo: la triangolazione (figura 3.17).

Conoscendo i parametri intrinseci ed estrinseci del sensore è possibile ricostruire la scena 3D.

Conoscendo le coordinate di 2 punti coniugati, generati dalla stesso punto 3D, le coordinate di 2 punti coniugati, generati dalla stesso punto 3D $m = [u, v, 1]$ e $m' = [u', v', 1]$ e le matrici di proiezione prospettica MPP delle due fotocamere $m \simeq PM$ e $m' \simeq P'M$ sappiamo che vale la relazione 3.23 per il punto m .

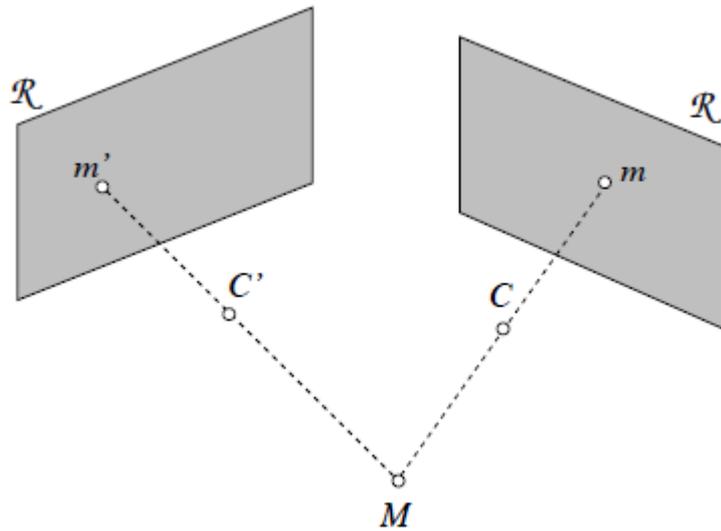


Figura 3.17. fig:Schema trinangolazione 3D

$$\begin{cases} u = \frac{p_1^T M}{p_3^T M} \\ v = \frac{p_2^T M}{p_3^T M} \end{cases} \quad (3.23)$$

Eseguendo alcuni passaggi matematici e considerando anche le equazioni per il punto coniugato otteniamo un sistema lineare omogeneo di quattro equazioni in quattro incognite.

$$\begin{bmatrix} (p_1 - up_3)^T \\ (p_2 - vp_3)^T \\ (p'_1 - u'p'_3)^T \\ (p'_2 - v'p'_3)^T \end{bmatrix} M = 0_{4 \times 1} \quad (3.24)$$

La soluzione è il nucleo della matrice 4×4 dei coefficienti, che dunque deve possedere rango tre. altrimenti incorreremmo nella soluzione banale $M = 0$. Inoltre in presenza di rumore la condizione sul rango non viene soddisfatta e la soluzione va ricercata attraverso la decomposizione ai valori singolari (SVD) Hartley e Sturm chiamano questo metodo linear-eigen.

Quanto esposto si generalizza nel caso il numero di fotocamere sia $N > 2$, ogni fotocamera aggiunge due equazioni e si ottiene un sistema omogeneo di $2N$ equazioni in quattro incognite.

Adesso consideriamo la situazione in cui alcuni punti di una scena siano proiet-

tati in due differenti fotocamere e per ogni punto della fotocamera sia dato il punto corrispondente nell'altra. Andiamo in contro al problema della posa relativa, ovvero la determinazione della posizione e dell'orientamento di una fotocamera rispetto all'altra, assumendo noti i parametri interni.

Possiamo anche parlare di recupero del moto della fotocamera, assumendo implicitamente lo scenario equivalente al precedente, in cui la scena statica viene inquadrata da una fotocamera in movimento con parametri interni noti. Nota l'informazione sulla posa relativa o equivalentemente sul moto, il recupero della struttura avviene mediante triangolazione, come nel caso calibrato, tale processo è noto con il nome di *structure from motion*.

La soluzione a tale problema, proposta da Longuet-Higgins, è ottenuta calcolando la *matrice essenziale*, che descrive la geometria epipolare di due immagini prospettiche con parametri interni noti e dalla quale è possibile ricavare il moto rigido della fotocamera sfruttando la fattorizzazione della matrice.

Notiamo che la componente di traslazione dello spostamento è calcolabile solo a meno di un fattore di scala, poiché non è possibile, senza informazioni aggiuntive, determinare se il moto osservato nell'immagine è di un oggetto vicino alla fotocamera che si muove lentamente o di un oggetto lontano dalla fotocamera che si muove più velocemente.

Quanto detto ci permette di determinare la matrice essenziale, infatti supposto di avere una fotocamera, con parametri interni noti, che si muova in un ambiente statico seguendo una traiettoria sconosciuta, consideriamo due immagini prese dalla fotocamera in due istanti di tempo diversi e , per ipotesi, consideriamo che siano presenti nelle due immagini dei punti corrispondenti. Note le MPP della fotocamera nei due istanti di tempo P , P' e le coordinate normalizzate dei due punti corrispondenti nelle immagini $p = K^{-1}m$, $p' = K'^{-1}m'$, se prendiamo come riferimento mondo il sistema di riferimento della prima fotocamera valgono le relazioni 3.25.

$$P [I/0] \quad P' [I/0] G = [R/t] \quad (3.25)$$

Sostituendo queste equazioni nell'equazione di Longuet-Higgins 3.19 otteniamo la forma bilineare che lega i punti coniugati 3.26.

$$p'^T [t]_X R p = 0 \quad (3.26)$$

La matrice essenziale 3.28 contiene i coefficienti della forma, e dipende dalla rotazione e dalla traslazione.

$$E = [t]_X R \quad (3.27)$$

Tale matrice, quindi, presenta solo cinque gradi di libertà determinati dai tre parametri della rotazione e dai due parametri della traslazione, a meno di un fattore di scala. Per risolvere il problema di ambiguità profondità-velocità, ovvero il problema di non essere in grado di ricavare la scala assoluta senza la conoscenza di un ulteriore parametro, è necessario notare che la matrice 3.28 è singolare, avendo $\det [t]_X = 0$, quindi abbiamo bisogno di altri due vincoli. Per risolvere il problema sopraesposto ricorriamo al teorema di Huang e Faugeras che ci permette di fattorizzare la matrice essenziale in rotazione e traslazione.

Teorema: Una matrice E 3×3 può essere fattorizzata come prodotto di una matrice non nulla antisimmetrica e di una matrice di rotazione se e soltanto se E ha due valori singolari uguali non nulli ed un valore singolare uguale a zero.

La dimostrazione del teorema si ottiene definendo la SVD di E come $E = UDV^T$ dove $D = \text{diag}(1, 1, 0)$ e U e V ortogonali. Avremo che:

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{S'} \underbrace{\begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{R'} \quad (3.28)$$

dove S' è antisimmetrica e R' è la matrice di rotazione, quindi

$$E = US'R'V^T = (US'U^T) (UR'V^T) \simeq \underbrace{(US'U^T)}_S \underbrace{\det(UV^T)}_R (UR'V^T) \quad (3.29)$$

La fattorizzazione cercata è $E = SR$, infatti la matrice $US'U^T$ è antisimmetrica e la matrice $UR'V^T$ è di rotazione.

Procedendo in senso inverso, definiamo $E = SR$, dove R è matrice di rotazione e S è antisimmetrica, e consideriamo la seguente proposizione:

Proposizione: Sia A una matrice $2n \times 2n$ reale e antisimmetrica con autovalori $\pm\lambda_1, \dots, \pm\lambda_n$. allora esiste una matrice diagonale a blocchi $\Lambda = \text{diag}(\lambda_1 Z, \dots, \lambda_n Z)$ con $Z = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ ed una matrice ortogonale S $2n \times 2n$ tali che:

$$S^T A S = \Lambda \quad (3.30)$$

Quindi vale la relazione $S \simeq U S' U^T$, e dall'equazione 3.28 si ottiene $S' = \text{diag}(1, 1, 0) R'^T$, dunque:

$$E \simeq U \text{diag}(1, 1, 0) (R'^T U^T R) \quad (3.31)$$

Poiché $R'^T U^T R$ è ortogonale, la 3.31 è la decomposizione ai valori singolari di E , con due valori singolari non nulli ed uno nullo, come volevasi dimostrare. L'aver dimostrato che la matrice essenziale è scomponibile tramite i suoi valori singolari ci torna utile per il calcolo della matrice stessa, ai fini della modellazione 3D.

Infatti considerando un insieme di corrispondenze di punti, tra due immagini, vogliamo determinare la matrice essenziale E che collega i punti secondo la relazione bilineare 3.32.

$$p_i'^T E p_i = 0 \quad (3.32)$$

La matrice E è calcolata sfruttando la vettorizzazione di un matrice (vedi appendice A.3) e il prodotto di Kronecker ((vedi appendice A.3)), ottenendo la relazione 3.33.

$$p_i'^T E p_i = 0 \Leftrightarrow \text{vec}(p_i'^T E p_i) = 0 \Leftrightarrow (p_i^T \otimes p_i'^T) \text{vec}(E) = 0 \quad (3.33)$$

Da cui si nota che ogni corrispondenza di punti genera un'equazione omogenea lineare di nove elementi incogniti della matrice E , inoltre da n punti corrispondenti otteniamo un sistema lineare di n equazioni 3.34.

$$\begin{bmatrix} p_1^T \otimes p_1'^T \\ p_2^T \otimes p_2'^T \\ \vdots \\ p_n^T \otimes p_n'^T \end{bmatrix} \text{vec}(E) = 0 \quad (3.34)$$

la soluzione al sistema lineare omogeneo è il nucleo di U_n , in particolare con $n = 8$ il nucleo della matrice ha dimensione uno e quindi la soluzione è determinata a meno di una costante moltiplicativa. Tale metodo prende il nome di algoritmo degli otto punti e rappresenta una variante del metodo Direct Linear transform (DLT)(vedi appendice A.6).

In pratica è possibile ottenere gli elementi di E risolvendo un problema lineare ai minimi quadrati e la soluzione è data dall'autovettore unitario corrispondente al minimo autovettore di $U_n^T U_n$, calcolabile col la SVD di U_n . Per concludere la trattazione ricordiamo che la procedura da seguire per il procedimento SfM, con ingresso i punti corrispondenti nelle immagini e i parametri interni K , è data dai seguenti passi:

1. Calcolo della matrice E con il metodo degli otto punti.
2. Fattorizzazione di E in $S = [T]_X R$
3. Definizione di due MPP $P = [I/0]$ e $P' = [R/t]$ ed effettuare al triangolazione dei punti ottenuti.

L'uscita del nostro processo è data dalla struttura del sistema, ovvero coordinate dei punti, e il moto R, t .

Capitolo 4

Applicativi per la valorizzazione dei beni culturali

4.1 VisualSfM

Dopo aver illustrato la teoria che costituisce le fondamenta della visione tridimensionale passiamo alla fase pratica che sfruttando quanto detto permette di trasformare in oggetti multimediali 3D. I metodi di Image based-modeling, ai quali appartiene la fotogrammetria, utilizzano la corrispondenza che esiste tra i punti omologhi di due o più immagini per ricostruire l'informazione tridimensionale della superficie di un oggetto mediante un preciso modello matematico. Tale tecnica unita alla metodologia detta Structure From Motion (SfM) permette di ottenere ottimi risultati nell'ambito della ricostruzione 3D mantenendo bassi costi e complessità computazionale non troppo elevata. Nella SfM considerate diverse viste di una scena ottenute mediante una fotocamera in movimento e dato un insieme di punti corrispondenti, è possibile ricostruire il moto della fotocamera, siccome i parametri intrinseci sono noti, conoscendo anche i parametri estrinseci, il moto rigido della fotocamera, grazie alla matrice essenziale, si ha la completa conoscenza delle MPP delle fotocamere e quindi mediante la fase di triangolazione è possibile ricavare la struttura della scena. Il primo applicativo ampiamente studiato dal punto di vista matematico e di utilizzo è VisualSfM. In realtà è una interfaccia grafica progettata da Changchang Wu[1] che ricostruisce la pipeline per ottenere una sparse reconstruction mettendo insieme tutti gli algoritmi necessari a tale scopo Sift e Bundle Adjustment. In più, integra gli algoritmi sviluppati da Furu-

kawa, Patch-based Multi-view Stereo (PMVS)[2] e CMVS(Clustering Views for Multi-view Stereo)[3] per il calcolo delle dense reconstruction attraverso la scomposizione del problema in cluster.

Lo schema a blocchi, in figura 4.1, sintetizza i passi necessari per la rappresentazione 3D di una scena reale, a partire da un set di immagini, il cui numero deve essere abbastanza elevato.

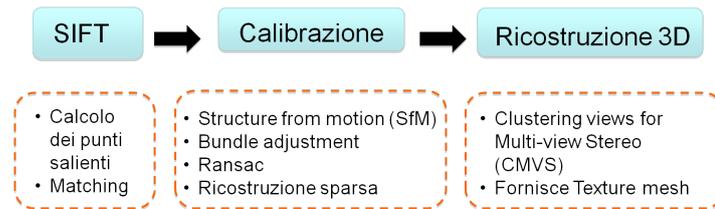


Figura 4.1. Schema a blocchi per la modellazione 3D in VisualSFM

L'applicativo innanzitutto calcola i punti salienti, per ciascuna immagine del set, utilizzando l'algoritmo SIFT (Scale-invariant feature transform), sviluppato da David Lowe, che estrae automaticamente da un'immagine le features, ovvero bordi, angoli e punti con particolare una tessitura facilmente individuabili e distinti rispetto agli altri all'interno di un'altra immagine. Il SIFT fornisce, per ogni punto saliente presente in un'immagine, un descrittore che ha il compito di caratterizzare il punto stesso, questo è utilizzato per l'operazione di matching tra coppie di immagini, cioè per la ricerca di corrispondenze tra tutte le coppie di immagini presenti nel nostro set, sfruttando la distanza euclidea.

Per la determinazione della matrice fondamentale è necessario utilizzare l'algoritmo otto-punti DLT (vedi appendice A.6), e data la ridondanza dei dati, anche l'algoritmo RANSAC (RANdom SAMple Consensus)[4], un metodo di ottimizzazione iterativo per la stima dei parametri di un modello matematico a partire da un insieme di dati contenente outlier. È un algoritmo non deterministico ovvero produce un risultato corretto solo con una data probabilità, che aumenta al crescere delle iterazioni che si effettuano.

La seconda fase prevede la calibrazione, il calcolo della matrici prospettiche, ovvero dei parametri intrinseci ed estrinseci della fotocamera, ed infine la ricostruzione 3D sparsa, ovvero costituita da una nuvola di punti. In particolare dopo aver trovato tutte le corrispondenze e le relative tracce di proiezione è necessario validare che esse risultino coerenti, poiché la presenza di molte

fotocamere ha come conseguenza l'accumulazione degli errori, che porta alla deriva del risultato, infatti gli errori sui parametri stimati sulle prime fotocamere si propaga quando vengono successivamente aggiunti altri elementi al modello. Per risolvere questo problema si deve cercare di minimizzare l'errore globale di proiezione di tutti i punti 3D nelle immagini, tale procedura, derivata dalla fotogrammetria, prende il nome di bundle adjustment. In pratica, date *fotocamere* e n punti 3D si cerca di aggiustare entrambi affinché la somma delle distanze euclidee al quadrato tra il punto j –esimo riproiettato tramite la fotocamera i –esima $P_i M^j$ ed il punto m_i^j sia la più piccola possibile:

$$\min_{R_i, t_i, M^j} \sum_{i=1}^N \sum_{j=1}^n d \left(K_i [R_i | t_i] M^j, m_i^j \right)^2 \quad (4.1)$$

La seconda fase prevede la calibrazione, il calcolo delle matrici prospettiche, ovvero dei parametri intrinseci ed estrinseci della fotocamera, ed infine la ricostruzione 3D sparsa, ovvero costituita da una nuvola di punti. In particolare dopo aver trovato tutte le corrispondenze e le relative tracce di proiezione è necessario validare che esse risultino coerenti, poiché la presenza di molte fotocamere ha come conseguenza l'accumulazione degli errori, che porta alla deriva del risultato, infatti gli errori sui parametri stimati sulle prime fotocamere si propaga quando vengono successivamente aggiunti altri elementi al modello. Per risolvere questo problema si deve cercare di minimizzare l'errore globale di proiezione di tutti i punti 3D nelle immagini, tale procedura, derivata dalla fotogrammetria, prende il nome di bundle adjustment. In pratica, date *fotocamere* e n punti 3D si cerca di aggiustare entrambi affinché la somma delle distanze euclidee al quadrato tra il punto j –esimo riproiettato tramite la fotocamera i –esima $P_i M^j$ ed il punto m_i^j sia la più piccola possibile:

$$\min_{R_i, t_i, M^j} \sum_{i=1}^N \sum_{j=1}^n d \left(K_i [R_i | t_i] M^j, m_i^j \right)^2 \quad (4.2)$$

dove K_i è la matrice dei parametri intrinseci, mentre R e t sono rispettivamente la matrice di rotazione e traslazione.

La minimizzazione della 4.2 equivale alla soluzione ai minimi quadrati di un sistema di equazioni non-lineari, risolto con l'algoritmo di Levenberg-Marquardt [5]. A questo punto il sistema sceglie la coppia di immagini che ha il maggior numero di corrispondenze e genera, mediante triangolazione, i

primi punti 3D relativi alle corrispondenze. Successivamente aggiunge un'altra telecamera al metodo di ottimizzazione, scelta in base ai punti 3D già presenti e dopo averne stimato i parametri estrinseci utilizza l'algoritmo DLT (appendice A.6) all'interno della procedura RANSAC. La procedura è ripetuta aggiungendo una telecamera alla volta finché per ognuna è stato possibile triangolare il punto 3D delle relative corrispondenze.

Poiché il bundler fornisce, in uscita, una ricostruzione 3D costituita da una nuvola di punti è necessario ricorrere al software (PMVS), sviluppato da Yasutaka Furukawa e Jean Ponce ([2]), per ottenere una ricostruzione 3D densa. Nelle figure sono riportati i risultati della ricostruzione 3D.

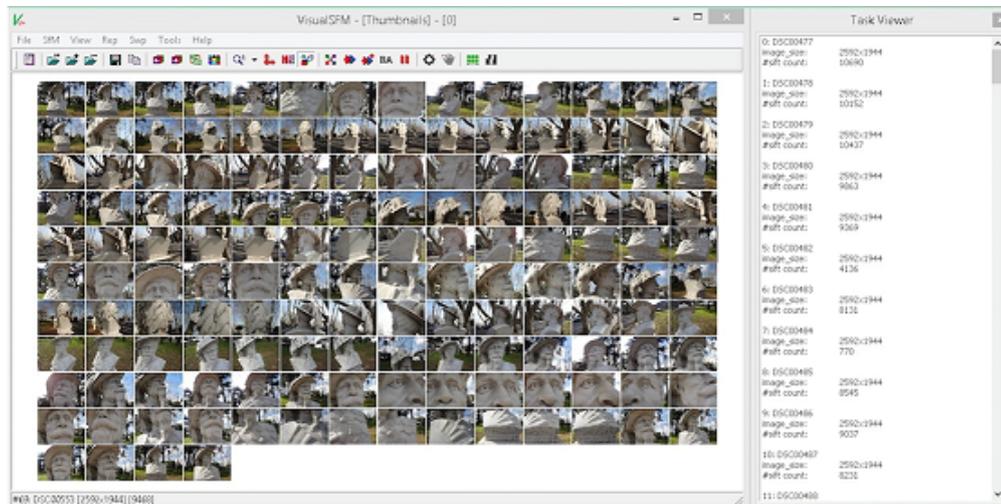


Figura 4.2. Acquisizione delle immagini e calcolo del SIFT

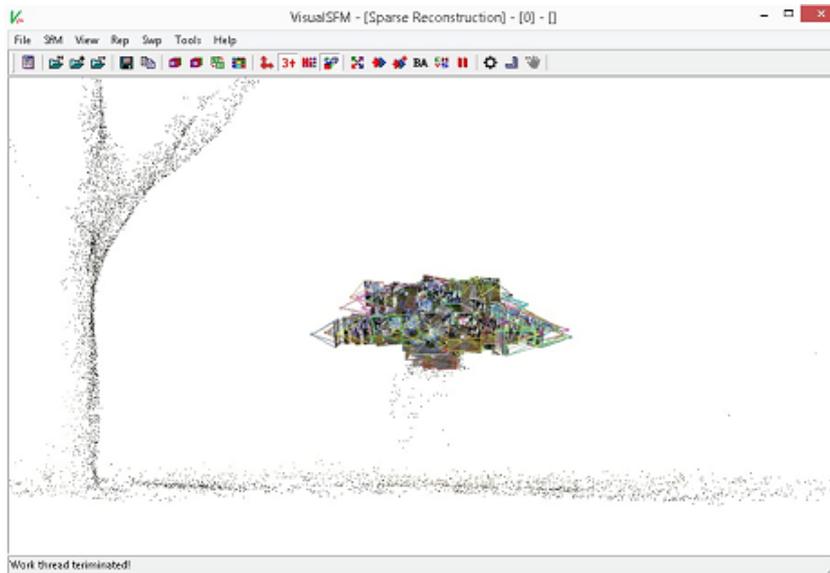


Figura 4.3. Ricostruzione sparsa



Figura 4.4. Ricostruzione 3D visualizzata con Meshlab

4.2 Panotour

Il secondo applicativo, utilizzato per la realizzazione di tour virtuali, è Panotour. Il lavoro sviluppato si inserisce nel progetto di valorizzazione del Colle Aventino, a Roma, nato dalla collaborazione tra la Fondazione Ugo Bordoni e la Sovrintendenza Speciale per i Beni Archeologici di Roma, che oltre a fornire fondamentali informazioni riguardo i luoghi d'interesse, ha concesso l'opportunità di accedere ai siti archeologici.

Nel corso del progetto sono stati sviluppati tre tour virtuali relativi a tre diversi siti archeologici, utilizzati, come vedremo nel prossimo paragrafo, anche all'interno dell'applicazione android:

- Casa Bellezza: è un importante esempio di edilizia tardo repubblicana di elevato livello sociale. Le strutture dell'abitazione si trovano attualmente a 12 metri di profondità e consistono di due stanze affrescate, realizzate in opus reticulatum con inserti in laterizio, che si aprono su un criptoportico in opus concretum, più un terzo ambiente tuttora interrato anch'esso probabilmente dipinto.
- Il Mitreo di Santa Prisca: è un santuario con una pianta rettangolare con copertura della volta a botte; è preceduto da due nicchie nelle quali erano collocate in origine due statue policrome dei portatori di fiaccola: Cautes e Cautopates. Nella nicchia di fondo si trova il caratteristico speleum, la grotta dove il dio sarebbe nato.
- La Domus di Via Marcella: sono resti di ambienti ipogei antichi posti nel piano interrato di villino moderno e in parte sotto il giardino, a circa 5 metri di profondità dal livello stradale, si compongono di cinque ambienti orientati secondo un asse nord-sud e realizzati con la medesima tecnica muraria, l'opera laterizia.

L'applicativo nasce come evoluzione del programma Krpano, ideato, sviluppato e gestito dal programmatore austriaco Klaus Reinfield ma per Krpano non esiste una vera interaccia ed il tour va creato e gestito tramite linguaggio xml. Tutti gli aspetti del virtual tour, infatti, devono essere implementati tramite file xml e html, per questo motivo la Kolor, con l'aiuto dello stesso Reinfield, ha creato Panotour come interfaccia friendly di Krpano.

Panotour permette di inserire e gestire panoramiche sferiche totali e parziali, e

garantisce la conversione da immagini sferiche a immagini cubiche e viceversa. Permette inoltre di generare transizioni, effetti dinamici, animazioni attraverso il collegamento di più panorami, contestualizzando il tuo virtuale grazie all'inserimento di mappe, loghi, suoni, immagini video e testi.

Lo schema a blocchi mostrato in figura 4.7 rappresenta le cinque componenti minime indispensabili per la realizzazione di un tour virtuale attraverso l'utilizzo del software Panotour, a queste verranno aggiunti, se necessario, altri file, come ad esempio audio, video e testi.

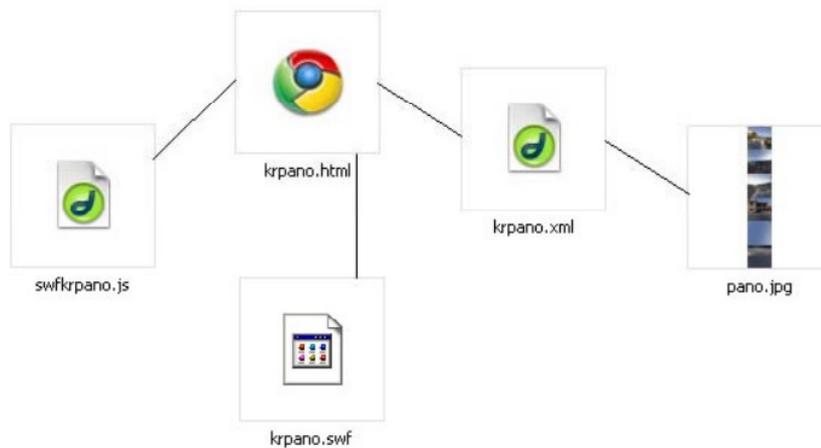


Figura 4.5. Schema interazione file in Krpano

L'intero tour virtuale è descritto e gestito dal file xml, che organizza logicamente gli elementi multimediali presenti nel tour stesso. Il file javascript esegue il parser del file xml gestendo le interazioni tra gli elementi, e per dispositivi mobili l'interazione con il giroscopio per una visita maggiormente interattiva. Il file flash player si occupa del render, ovvero permettere la visualizzazione della visita virtuale. Infine il file html permette al browser di interpretare e visualizzare il tour virtuale, e, al tempo stesso, fa da collante tra i numerosi file esistenti nella cartella di Panotour.

In pratica partendo da un insieme di immagini disposte su un supporto sferico è possibile tramite una trasformazione di coordinate disporre le immagini in un cubo, indispensabile per la costruzione della visita virtuale.



Figura 4.6. Immagini disposte sulle facce di un cubo



Figura 4.7. Tour virtuale

4.3 Aventino: applicazione Android

Nell'ambito del progetto Rome ArcheoMedia è stata sviluppata una applicazione multi-piattaforma (Android, iOS) dedicata all'archeologia che consente al visitatore di avere una maggiore percezione dei luoghi e delle strutture storico archeologiche inserite nel paesaggio urbano, osservando i cambiamenti avvenute nelle diverse epoche storiche, e riuscendo tramite filmati, immagini e realtà virtuale, sul dispositivo mobile, a visitare anche luoghi invisibili come resti archeologici sotterranei non più visitabili e le testimonianze inserite nel contesto urbano e per questo poco comprensibili.

L'applicazione costituita da numerose pagine navigabili, permette di visitare i siti archeologici presenti nel Colle Aventino a Roma, e attraverso immagini, testi, audio e video, fruibili anche senza connessione ad internet o comunque mediante streaming adattativo alla banda disponibile, garantisce una visita completa e stimolante.

La potenzialità dell'applicazione risiede nella sua progettazione architettonica, che, grazie alle nozioni basilari dell'ingegneria del software, ha permesso di ottenere la massima modularità del codice. Infatti gli elementi multimediali, immagazzinati in un server, sono stati organizzati in maniera tale da permettere facilmente la creazione di altre applicazioni in modo semi-automatico, basta solo rinominare i file per la nuova applicazione.

Il cuore dell'applicazione risiede in un file xml contenente tutte le informazioni necessarie per la creazione dell'applicazione, ad esempio le informazioni relative ad un determinato sito archeologico: numero delle foto, le sue coordinate GPS, la presenza o meno di realtà virtuale. Riportiamo un estratto del file 4.8,

```
<?xml version="1.0" encoding="utf-8" ?>
<sites>
  <site id="01">
    <name>CASA BELLEZZA</name>
    <didascalia> </didascalia>
    <icon>00</icon>
    <lat>41.883876</lat>
    <longi>12.482765</longi>
    <numFoto>20</numFoto>
    <percorso>1,1,0</percorso>
    <ordinePercorsi>3,2,0</ordinePercorsi>
    <subsite>LA DOMUS,LA STANZA DELLE COLONNE,IL CRIPTOPORTICO,LA STANZA GIALLA</subsite>
    <numfotosubsite>2,8,2,8</numfotosubsite>
    <vr>1</vr>
    <fotovideo>1,1</fotovideo>
  </site>
</sites>
```

Figura 4.8. Sezione del file xml utilizzato per l'applicazione

Inoltre all'interno dell'applicazione, disponibile anche in lingua inglese, sfruttando le API messe a disposizione da Google, è possibile avviare la navigazione GPS verso il punto di interesse oppure scansionare i codici QRcode, posizionati nel percorso di visita, per aumentare l'interattività dell'utente. Mostriamo alcuni screenshot 4.9 della nostra applicazione.

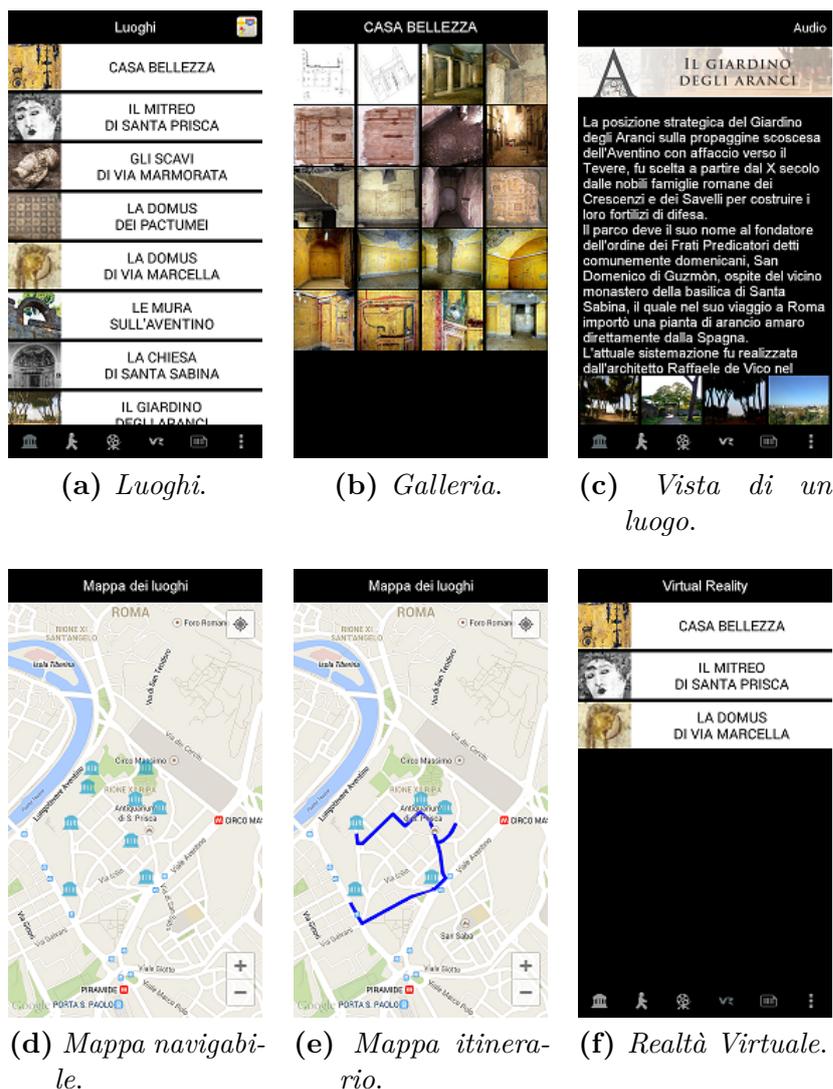


Figura 4.9. Screenshot dell'applicazione Aventino per dispositivi Android

Per concludere l'applicazione è stata testata sia su differenti dispositivi cellulari, con schermi di diversa dimensione e risoluzione, sia con utenti differenti, per valutarne la sua qualità oggettiva e soggettiva.

Capitolo 5

Conclusioni

Il lavoro svolto durante il periodo di dottorato ha rispettato gli obiettivi proposti all'atto della stesura del progetto, fornendo un contributo alla ricerca svolta presso la Fondazione Ugo Bordoni, dove ho trascorso gran parte del mio tempo. Nel periodo di studio è stato possibile realizzare degli strumenti matematici e software utili per la valorizzazione, la tutela e la corretta fruizione dei Beni Culturali, sfruttando da una parte la capacità di elaborazione dei computer desktop, per la realizzazione ricostruzioni 3D e simulazioni, dall'altra le potenzialità, sempre più sorprendenti, dei recentissimi smartphone, che oramai condizionano, in bene o in male, la nostra vita. In questi anni, non senza difficoltà, restando sempre al passo con la tecnologia e con le nuove apparecchiature, che hanno letteralmente invaso il mercato, il lavoro svolto ha permesso di mantenere il tema di ricerca fortemente attuale e sfruttando l'onda di interesse e modernizzazione che, in particolare in questo ultimo periodo, ha interessato l'ambito Beni Culturali, si proietta verso molteplici sviluppi futuri. Infatti il lavoro ha posto le basi ad alcune proposte di partecipazione al nuovo Programma Quadro europeo per il finanziamento della Ricerca e dell'Innovazione (2014 - 2020), denominato HORIZON 2020.

Inoltre nell'ambito della convenzione tra la Fondazione Ugo Bordoni e l'Istituto di Scienze e Tecnologie per la Cognizione del CNR sto partecipando allo sviluppo di ambienti digitali, ovvero di simulazioni al computer del contesto storico, sociale, economico, e politico in cui è inserito un bene archeologico (sito, monumento, bene conservato in un museo). In quanto simulazioni di meccanismi e processi di cause e effetti, gli ambienti digitali permettono agli utenti di capire il contesto di un bene archeologico modificando fattori e vedendo le

conseguenze di queste modifiche, assumendo ruoli da soli o con altri, e facendo confronti con altri contesti, del passato e di oggi. Gli ambienti digitali hanno un'interfaccia utente principalmente non verbale, e fortemente interattiva e motivante come un computer game. Possono essere molto semplificati o più realistici, possono esistere in diverse versioni per diversi tipi di utenti, e gli stessi utenti possono arricchire l'ambiente digitale con l'aggiunta di nuove informazioni. Personalmente, nel corso di questi anni trascorsi tra seminari, scuole dottorali, convegni, riunioni e quant'altro, posso felicemente affermare di essere cresciuto sia dal punto di vista umano, rapportandomi con persone di cultura e idiomi differenti, sia dal punto di vista lavorativo, affrontando i problemi in maniera alternativa e trovando le soluzioni attraverso una buona dose di tenacia e pazienza.

Appendice A

Strumenti matematici

A.1 Espansione di Laguerre Gauss

L'espansione in funzioni di Laguerre-Gauss di un'immagine è una particolare decomposizione in funzioni circolari armoniche (Circular Harmonic Functions) di un segnale bidimensionale e può essere considerata una forma generalizzata dello sviluppo in serie di Taylor in forma polare se si interpretano le coordinate come numeri complessi.

La luminanza di un'immagine è rappresentata nel dominio bidimensionale reale e continuo di coordinate x e y come una funzione reale $f(x, y)$ nel dominio dell'immagine, la rappresentazione di $f \in L^2(\mathfrak{R}^2, d^2x)$ in coordinate polari (r, γ) attorno al punto ξ è data da:

$$f_p(r, \gamma) = f(\xi_1 + r \cos(\gamma), \xi_2 + r \sin(\gamma)). \quad (\text{A.1})$$

La funzione $f_p(r, \gamma)$ è periodica rispetto a γ con periodo 2π e può essere espansa in serie di Fourier ovvero decomposta in una somma di funzioni circolari armoniche $\psi^{(n)}[x(r, \gamma), y(r, \gamma)] = h_n(r) e^{jn\gamma}$ come

$$f_p(r, \gamma) = \sum_n h_n(r) e^{jn\gamma} \quad (\text{A.2})$$

dove $h_n(r)$ è il profilo radiale relativo all'armonica circolare di ordine n ed è definito dall'espressione:

$$h_n(r) = \frac{1}{2\pi} \int_0^{2\pi} f_p(r, \theta) e^{-jn\theta} d\theta. \quad (\text{A.3})$$

Ciascuna CHF è polarmente separabile sia nel dominio dello spazio che in quello della frequenza, infatti la sua trasformata di Fourier $\Psi^{(n)}(\rho, \vartheta)$ è pari a

$$\Psi^{(n)}(\rho, \vartheta) = (-j)^n \tilde{h}_n(\rho) e^{jn\vartheta} \quad (\text{A.4})$$

con ρ e ϑ coordinate polari nel dominio della frequenza e $\tilde{h}_n(\rho)$ trasformata di Hankel di ordine n di $h_n(r)$:

$$\tilde{h}_n(\rho) = 2\pi \int_0^{+\infty} r h_n(r) J_n(r\rho) dr \quad n = 0, 1, 2, \dots \quad (\text{A.5})$$

Ogni profilo radiale $h_n(r)$ può essere sviluppato rispetto ad un insieme di funzioni ortogonali sul dominio dell'immagine, in particolare se si considerano i polinomi generalizzati di Laguerre $L_m^{(\alpha)}(x)$, definiti dalla relazione differenziale

$$L_m^{(\alpha)}(x) = \frac{1}{m! e^{-x} x^\alpha} \frac{d^m}{dx^m} \{ e^{-x} x^{\alpha+m} \} \quad (\text{A.6})$$

si ottiene la decomposizione dell'immagine $f_p(r, \gamma)$ in funzioni circolari armoniche di Laguerre-Gauss $L_k^{(n)}(r, \gamma)$

$$f_p(r, \gamma) = \sum_n \sum_k y_{n,k} L_k^{(n)}(r, \gamma) \quad (\text{A.7})$$

con

$$L_k^{(n)}(r, \gamma; \sigma) = \frac{(-1)^k}{\sqrt{k! (|n| + k)!}} \left(\frac{r}{\sigma} \right)^{|n|} e^{jn\gamma} * L_k^{(|n|)} \left[\left(\frac{r}{\sigma} \right)^2 \right] \frac{1}{\sigma \sqrt{\pi}} e^{-\frac{1}{2} \left(\frac{r}{\sigma} \right)^2} \quad (\text{A.8})$$

In figura A.1 è mostrato il profilo delle funzioni armoniche circolari di Laguerre-Gauss al variare degli indici n e k .

Nella A.8 il polinomio generalizzato di Laguerre $L_k^{(n)}(2\pi r^2)$ può essere calcolato con la formula di Rodrigues:

$$L_k^{(n)} = \sum_{h=0}^k (-1)^h \binom{n+k}{k-h} \frac{x^h}{h!} \quad (\text{A.9})$$

Dalla A.8 è immediato dedurre che la parte reale e quella immaginaria della $L_k^{(n)}(r, \theta)$ hanno nello spazio andamenti identici a meno di una rotazione di 90° come è mostrato nelle figure ?? e ??. Dall'ispezione dell'andamento dei

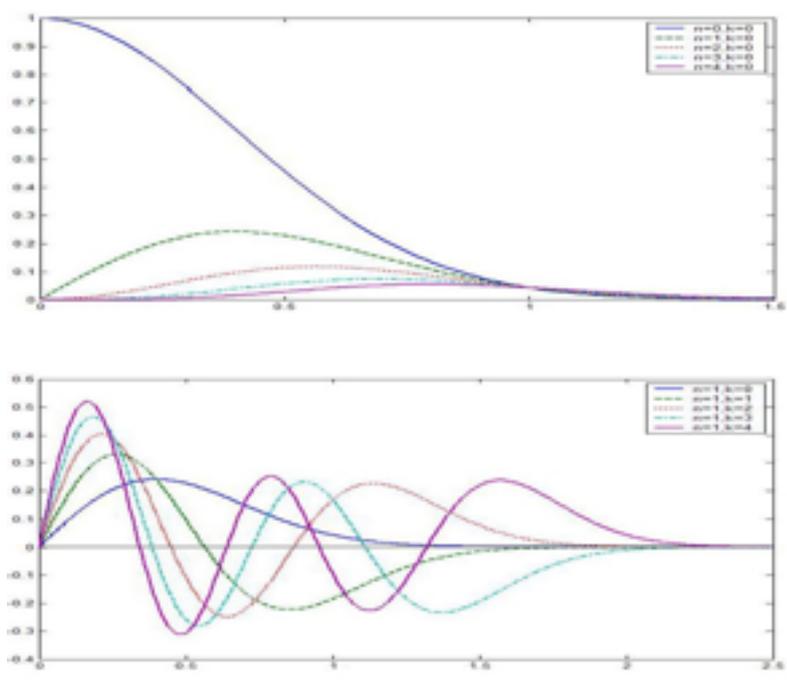


Figura A.1. Profilo radiale delle CHF di LG rispetto a n e k

profili di risulta evidente come il numero di oscillazioni radiali della funzione di LG di ordine (n, k) sia pari a $(k + 1)$.

Una proprietà delle CHF $L_k^{(n)}(r, \theta)$ è quella di essere autofunzioni della trasformata di Hankel di ordine k , conseguentemente risultano angolarmente e radialmente isomorfe con il loro spettro di Fourier dato da:

$$\hat{L}_k^{(n)}(\rho, \gamma) = (-1)^k (-j)^n \left(\frac{\rho}{2\pi}\right)^n L_k^{(n)} \left[(2\pi)^{-1} \rho^2 \right] e^{-\frac{\rho^2}{4\pi}} e^{jn\gamma} \quad (\text{A.10})$$

Le CHF di Laguerre-Gauss costituiscono un sistema di funzioni ortogonali completo per le funzioni $f \in L^2(\mathbb{R}^2)$ per cui i coefficienti $y_{n,k}$ sono dati dal prodotto scalare tra l'immagine e gli elementi della base: $y_{n,k} = \langle f_p(r, \theta), L_k^n(r, \theta) \rangle$. Questo prodotto scalare equivale a calcolare la convoluzione di $f(x, y)$ per la $L_k^{(n)}[r(x, y), \theta(x, y)]$ nel punto ξ ovvero a filtrare l'immagine con $\hat{L}_k^{(n)}(\rho, \gamma)$. Le funzioni di LG per la loro particolare dipendenza dalla coordinate angolari sia nel dominio dello spazio che in quello della frequenza sono self-steerable: ciò significa che possono essere orientate in qualsiasi direzione semplicemente moltiplicando la CHF per il numero complesso $e^{-jn\varphi}$. Se si indica con R_φ l'operatore di rotazione per la self-steerability di $L_k^{(n)}(r, \gamma)$ si può scrivere:

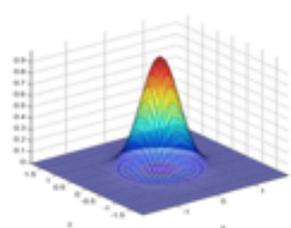
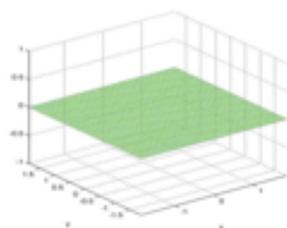
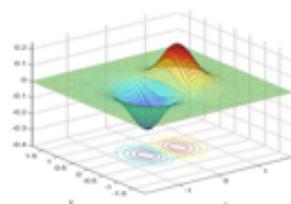
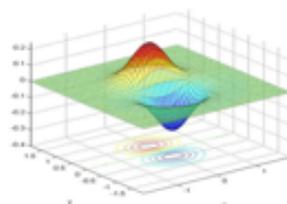
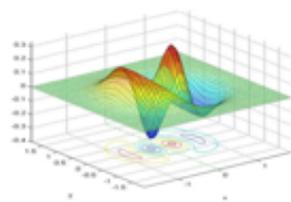
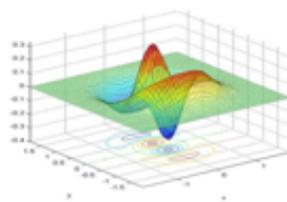
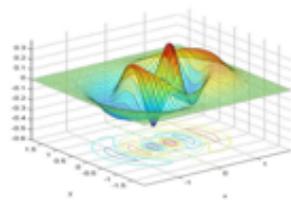
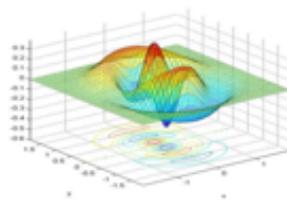
(a) *Reale* $n=0, k=0$.(b) *Immaginaria* $n=0, k=0$.(c) *Reale* $n=1, k=0$.(d) *Immaginaria* $n=1, k=0$.(e) *Reale* $n=1, k=1$.(f) *Immaginaria* $n=1, k=1$.(g) *Reale* $n=1, k=2$.(h) *Immaginaria* $n=1, k=2$.

Figura A.2. Andamenti nello spazio delle funzioni di LG definite per $n = 0, k = 0$ e $n = 1, k = 0, 1, 2$

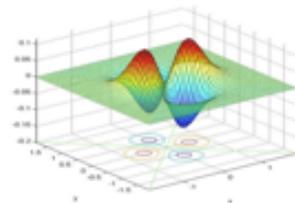
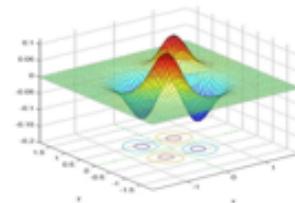
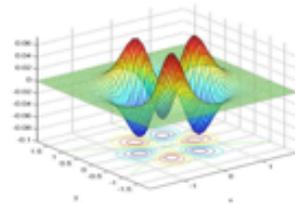
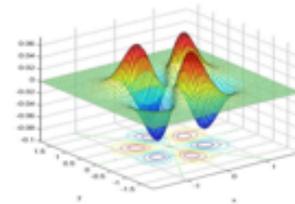
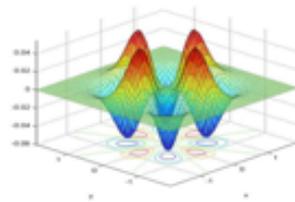
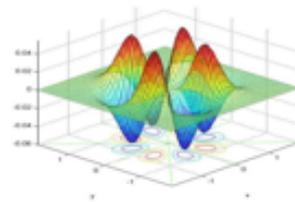
(a) *Reale* $n=2$, $k=0$.(b) *Immaginaria* $n=2$, $k=0$.(c) *Reale* $n=3$, $k=0$.(d) *Immaginaria* $n=3$, $k=0$.(e) *Reale* $n=4$, $k=0$.(f) *Immaginaria* $n=4$, $k=0$.

Figura A.3. Andamenti nello spazio delle funzioni di LG definite per $n = 2, 3, 4, k = 0$

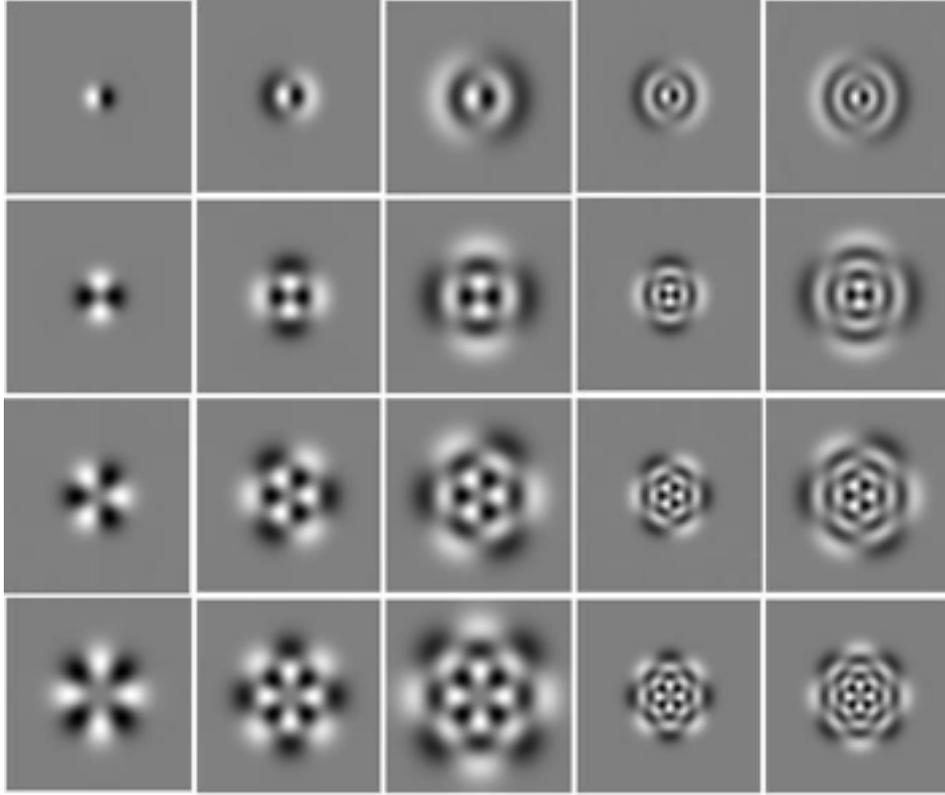


Figura A.4. Andamento in frequenza della parte reale dei filtri di LG per $n = 1..4$ (da sinistra a destra) e $k = 1..4$ dall'alto verso il basso

$$R_\varphi [L_k^{(n)}(r, \gamma)] = e^{-jn\varphi} L_k^{(n)}(r, \gamma) \quad (\text{A.11})$$

da cui

$$R_\varphi [f_p(r, \gamma)] = R_\varphi \left[\sum_n \sum_k y_{n,k} L_k^{(n)}(r, \gamma) \right] = \sum_n \sum_k (e^{-jn\varphi} y_{n,k}) L_k^{(n)}(r, \gamma) \quad (\text{A.12})$$

La A.12 può essere interpretata dicendo che nel dominio trasformato la rotazione di una struttura attorno ad un fissato centro di rotazione corrisponde alla moltiplicazione per $e^{-jn\varphi}$ dei coefficienti complessi $y_{n,k}$ calcolati rispetto al centro di rotazione. Questa caratteristica rende lo sviluppo di Laguerre-Gauss particolarmente adatto all'analisi orientazioni locali delle immagini.

Il parametro σ regola l'ampiezza della gaussiana e consente di analizzare l'immagine su diverse scale di risoluzione A.4. Nella scelta di σ si deve tener conto dell'aliasing dovuto alle code delle forma d'onda dei filtri quando questi

sono progettati per campionamento in frequenza per effettuare i filtri avvalendosi degli algoritmi veloci di FFT. Per concentrare l'energia nella banda $[-\pi, \pi]$ è necessario aumentare la varianza σ nello spazio a scapito della risoluzione spaziale. 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.

Le tecniche di rappresentazione delle immagini attraverso i bordi 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. È possibile sfruttare l'informazione rappresentata dai bordi anche per la valutazione della qualità.

In particolare ci proponiamo di misurare la "coerenza dei bordi" rappresentando l'immagine localmente attraverso l'espansione di Laguerre-Gauss.

I filtri LG godono della proprietà di rilevare le diverse strutture presenti nell'immagine: bordi, intersezioni di linee, punti di vertice. Considerata un'immagine in coordinate polari $I(r, \gamma)$ con $r = \sqrt{x^2 + y^2}$ e $\gamma = \arctg\left(\frac{y}{x}\right)$, la sua espansione è espressa dalla relazione:

$$I(r, \gamma) = \sum_{n=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} y_{n,k} L_{n,k}(r, \gamma; \sigma). \quad (\text{A.13})$$

Il primo termine dell'espansione è calcolato come

$$y_{n,k}(\sigma) = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} I(x, y) L_{n,k}(x, y; \sigma) dx dy \quad (\text{A.14})$$

dove $L_{n,k}(x, y; \sigma)$ è la famiglia di Funzioni Circolari Armoniche di Laguerre Gauss. I coefficienti associati ai pixel dell'immagine, possono essere calcolati tramite la convoluzione

$$y_{n,k}(n_1, n_2, \sigma) = I(x, y) * L_{n,k}(-x, -y; \sigma) \quad (\text{A.15})$$

ricostituibile ad una semplice operazione di filtraggio. Il modulo dell'uscita del filtro $n = 1, k = 0$ è rappresentativo dell'intensità dei bordi rilevati nell'immagine A.5.

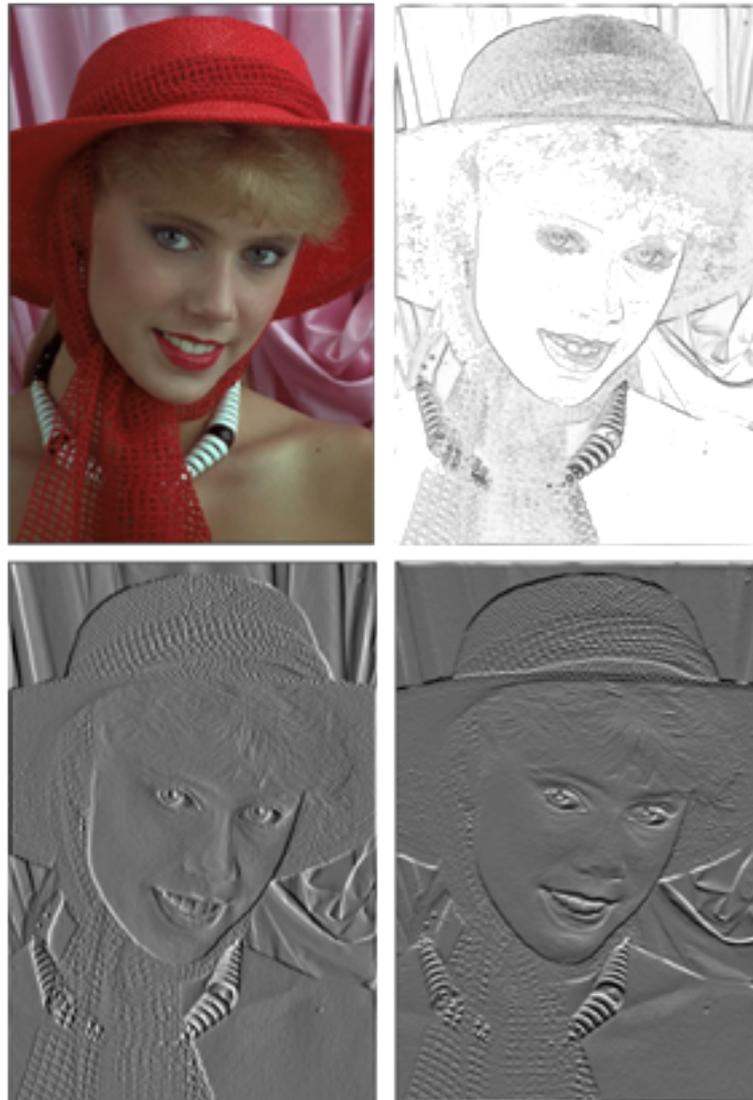


Figura A.5. Immagine womanhat (Database L.I.V.E. Release 2) filtrata mediante i filtri di Laguerre-Gauss $n = 1, k = 0$. Dall'immagine in alto a sinistra in senso orario: immagine originale, modulo di y_{10} , parte reale di y_{10} e parte immaginaria di y_{10}

Plesso per la valutazione della qualità

Acquistano notevole importanza, per applicazioni di image processing, un sottoinsieme di funzioni di Laguerre-Gauss, dette **Plesso** di Laguerre-Gauss. In particolare consideriamo le funzioni complesse di ordine angolare $n = 1, 3$ ed ordine radiale $k = 0$, utili per modellare il comportamento del sistema visivo umano. Il Plesso è definito da

$$L_{1,0}(r, \gamma; \sigma) = -\frac{1}{\sigma\sqrt{\pi}} e^{-\frac{r^2}{2\sigma^2}} \sqrt{\frac{r^2}{\sigma^2}} e^{j\gamma} \quad (\text{A.16})$$

$$L_{3,0}(r, \gamma; \sigma) = -\frac{1}{\sqrt{6}} \frac{1}{\sigma\sqrt{\pi}} e^{-\frac{r^2}{2\sigma^2}} \left[\sqrt{\frac{r^2}{\sigma^2}} \right]^3 e^{j3\gamma} \quad (\text{A.17})$$

Se consideriamo un'immagine $I(n_1, n_2)$ in coordinate polari, il plesso di LG assume la forma

$$h_1(n_1, n_2) = -\frac{1}{\sigma\sqrt{\pi}} e^{-\frac{n_1^2+n_2^2}{2\sigma^2}} \sqrt{\frac{n_1^2+n_2^2}{\sigma^2}} \cdot e^{j\arctg\left(\frac{n_2}{n_1}\right)} \quad (\text{A.18})$$

$$h_3(n_1, n_2) = -\frac{1}{\sqrt{6}} \frac{1}{\sigma\sqrt{\pi}} e^{-\frac{n_1^2+n_2^2}{2\sigma^2}} \left[\sqrt{\frac{n_1^2+n_2^2}{\sigma^2}} \right]^3 \cdot e^{j3\arctg\left(\frac{n_2}{n_1}\right)} \quad (\text{A.19})$$

con σ fattore di scala.

Notiamo che la risposta $h_1(n_1, n_2)$, filtro di LG con $n = 1, k = 0$, risulta essere è “accordato” sulla struttura dei bordi curvilinei, ossia si comporta come estrattore dei bordi dell'immagine A.6.

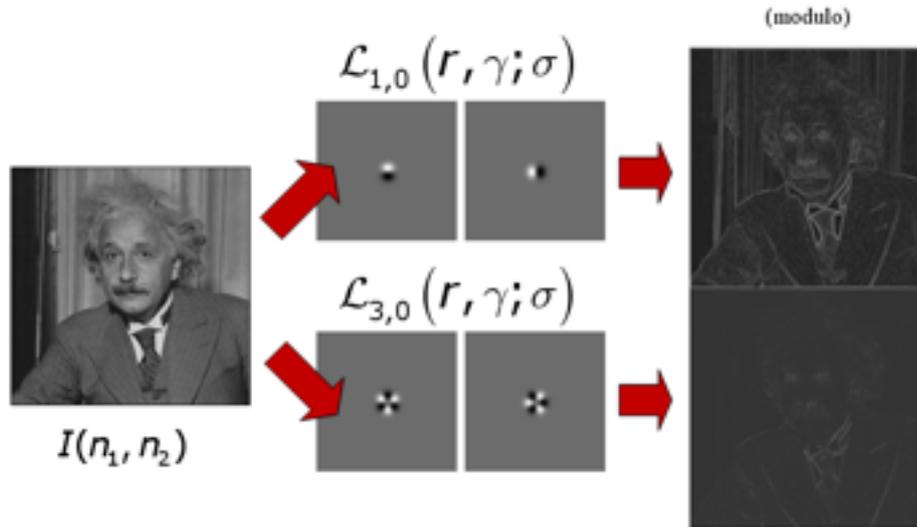


Figura A.6. Plesso dell'immagine Einstein

A.2 Circular Fitting

Per ottenere una maggiore regolarità di forme, e allo stesso tempo rispettare il principio di continuità e di direzione, l'interpolazione dei bordi nella zona danneggiata verrà effettuata tramite tecniche di circular fitting. La traiettoria è ottenuta tramite l'interpolazione di un cerchio ai minimi quadrati. Riconducendo il problema in un campo puramente geometrico.

In particolare se il raggio trovato supera una certa soglia verrà interpolata una retta attraverso i punti inseguiti.

A.2.1 Fitting Algebrico

Per definizione sappiamo che un cerchio è interpolato in modo algebrico ai minimi quadrati se i parametri a, b, c della sua equazione

$$F(\mathbf{x}) = a\mathbf{x}^T\mathbf{x} + \mathbf{b}^T\mathbf{x} + c = 0 \quad (\text{A.20})$$

sono determinati nel senso dei minimi quadrati con $a \neq 0$ e x, b reali.

A tale scopo consideriamo l'insieme dei punti i da interpolare ed inseriamoli nell'equazione A.20.

Si ottiene un sistema di equazioni lineari sovradimensionato che in genere non ha una soluzione, quindi il problema sarà risolvere un sistema in cui si

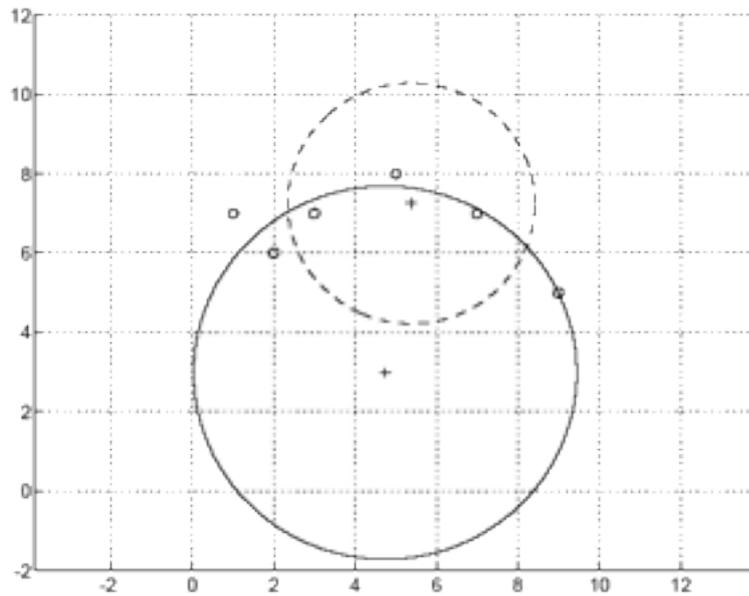


Figura A.7. Fitting algebrico (tratteggiato), Fitting geometrico non lineare (continuo)

cerca di minimizzare la quantità $B \cdot u = r$ dove $u = [a \ b_1 \ b_2 \ c]^T$ sono le incognite e $\mathbf{B} = \begin{pmatrix} a_{11}^2 + a_{12}^2 & a_{11} & a_{12} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}^2 + a_{m2}^2 & a_{m1} & a_{m2} & 1 \end{pmatrix}$ i punti noti e la soluzione tale da rendere il vettore r minimo. La soluzione A.7 non soddisfa i requisiti del sistema.

La soluzione algebrica può essere usata come prima approssimazione del fitting geometrico non lineare.

A.2.2 Fitting Geometrico non lineare

Dalla teoria è noto che un cerchio è interpolato in modo geometrico se si è minimizzata la quantità $\min \sum ||\mathbf{z} - \mathbf{x}_i|| - r|^2$, dove \mathbf{z} è il centro del cerchio, x_i sono i punti noti e r il raggio.

Tale sistema risulta non lineare che può essere risolto con il metodo iterativo di Gauss-Newton considerando come valore iniziale la soluzione del problema algebrico.

Il metodo di Newton, noto in letteratura, è un metodo classico per la risoluzione di equazioni attraverso approssimazioni successivamente. Il processo

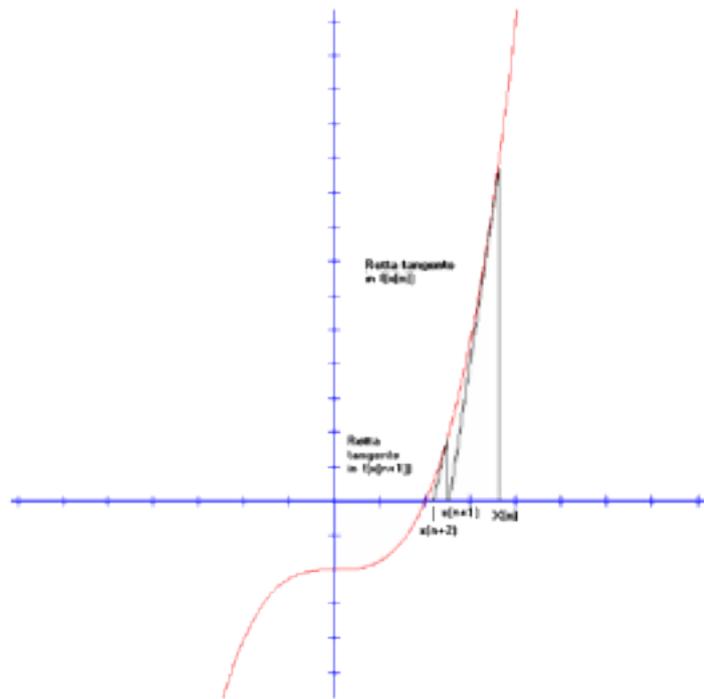


Figura A.8. Grafico della funzione iterativa di Gauss-Newton

iterativo tende a portare il sistema dinamico ad una soluzione stazionaria. Matematicamente e graficamente A.8 si ha:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad (\text{A.21})$$

In particolare, trattandosi di uno spazio bidimensionale, metodo di Gauss-Newton è applicato sfruttando l'operatore Jacobiano.

Definiamo $u = [z_1 \ z_2 \ r]^T$ vettore incognito e la funzione $f(\mathbf{u}) = (|\|\mathbf{z} - \mathbf{x}_i\| - r|)$, allora la soluzione del sistema $\sum_{i=1}^m f_i(\mathbf{u})^2 = \min$ è trovata tramite approssimazioni successive della soluzione $\tilde{\mathbf{u}} = \tilde{\mathbf{u}} + \mathbf{h}$ calcolando, di volta in volta, il fattore di correzione h .

Data la funzione f , consideriamo il suo sviluppo mediante la serie di Taylor al primo termine, otteniamo $f(\tilde{\mathbf{u}} + \mathbf{h}) \approx f(\tilde{\mathbf{u}}) + J(\tilde{\mathbf{u}}) \cdot \mathbf{h} \approx \mathbf{0}$ dove $J()$ indica il Jacobiano di f che sarà uguale

$$J(\mathbf{u}) = \begin{pmatrix} \frac{u_1 - x_{11}}{\sqrt{(u_1 - x_{11})^2 + (u_2 - x_{12})^2}} & \frac{u_2 - x_{12}}{\sqrt{(u_1 - x_{11})^2 + (u_2 - x_{12})^2}} & -1 \\ \vdots & \vdots & \vdots \\ \frac{u_m - x_{m1}}{\sqrt{(u_m - x_{m1})^2 + (u_m - x_{m2})^2}} & \frac{u_m - x_{m1}}{\sqrt{(u_m - x_{m1})^2 + (u_m - x_{m2})^2}} & -1 \end{pmatrix} \quad (\text{A.22})$$

Per trovare la soluzione al nostro problema è necessario, innanzitutto, calcolare il termine h attraverso la risoluzione del sistema lineare ai minimi quadrati $J(\tilde{\mathbf{u}}) \cdot \mathbf{h} \approx -f(\mathbf{u})$, da cui otteniamo $\mathbf{h} = -(\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T f(\mathbf{u})$.

Il passo successivo è aggiornare l'approssimazione $\tilde{\mathbf{u}} = \tilde{\mathbf{u}} + \mathbf{h}$, controllando, al tempo stesso, che venga soddisfatto il criterio di arresto $\|\mathbf{h}\| > 10^{-4} \|\tilde{\mathbf{u}}\|$.

La soluzione è mostrata in figura A.7.

Notiamo che per ottenere la soluzione desiderata è necessario utilizzare due metodi, ovvero il fitting algebrico nella fase iniziale del processo e il fitting geometrico per raggiungere l'ottimo. La combinazione dei due metodi consente di avere una ottima interpolazione ma purtroppo ha un elevato costo computazionale. La convergenza è fortemente influenzata dalla bontà della soluzione iniziale, cioè di quella algebrica.

Per ovviare al problema ricorriamo al metodo linear Least-Square.

A.2.3 Linear Least-Square

Il linear Least square è un metodo che minimizza non il quadrato dell'errore ma la quantità

$$\min \sum \left| (\mathbf{z} - \mathbf{x}_i)^2 - r^2 \right|^2. \quad (\text{A.23})$$

Per ottenere un sistema lineare imponiamo $z = \begin{bmatrix} z_1 & z_2 \end{bmatrix}^T$ le coordinate incognite del centro, r il raggio e $x_i = \begin{bmatrix} x_{i1} & x_{i2} \end{bmatrix}^T$ i punti osservati.

Possiamo riscrivere la sommatoria come $\mathbf{z}^T \mathbf{z} - 2\mathbf{z}^T \mathbf{x}_i + \mathbf{x}_i^T \mathbf{x}_i - r^2$ e definiamo $\mathbf{y} = \begin{pmatrix} 2z_1 \\ 2z_2 \\ r^2 - \mathbf{z}^T \mathbf{z} \end{pmatrix}$ e $b_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ 1 \end{pmatrix}$, ottenendo $\mathbf{b}_i^T \mathbf{y} = 2\mathbf{z}^T \mathbf{x}_i - \mathbf{z}^T \mathbf{z} + r$.

Effettuando un cambio di variabile, equivalentemente alla A.23, otteniamo

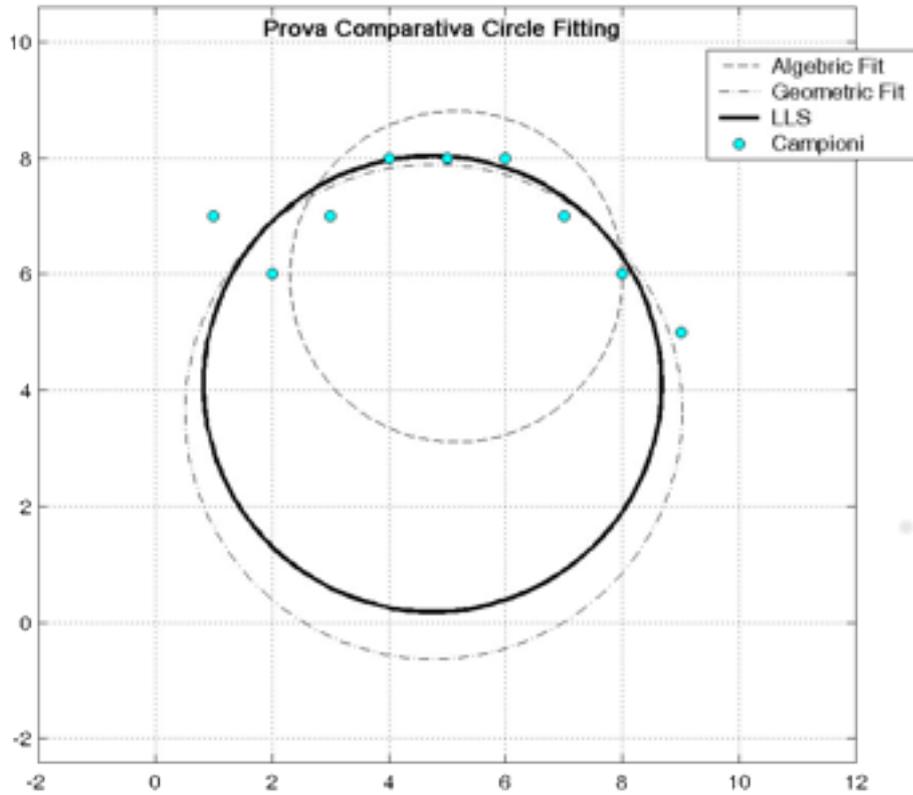


Figura A.9. Comparativa metodi di circular fitting: Algebrico, Geometrico, LLS

$$\min_y \sum_m (\mathbf{b}_i^T \mathbf{y} - \mathbf{x}_i^T \mathbf{x}_i)^2. \quad (\text{A.24})$$

Scritta in forma compatta $\min_y \|B\mathbf{y} - \mathbf{w}\|$, con $B = \begin{bmatrix} b_1 & b_2 & \dots & b_m \end{bmatrix}^T$ e \mathbf{w} è un vettore di componenti $\|\mathbf{x}_i\|^2$.

La soluzione del problema lineare è data da $\mathbf{y} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{W}$.

Graficamente è possibile comparare i tre metodi (figura A.9).

Il metodo LLS, grazie al suo costo computazionale contenuto, risulta il migliore, ma test di interpolazione tramite MatLab hanno mostrato che nel caso in cui punti siano vicini e posti su di una circonferenza con una curvatura leggera tale metodo perde la sua efficacia mentre converge a quello geometrico nel caso di piccole circonferenze.

In conclusione abbiamo scelto di utilizzare il metodo LLS per la prima fase del processo, sostituendo il metodo algebrico, e il metodo geometrico come sistema di interpolazione curvilinea dei bordi (figura A.10).

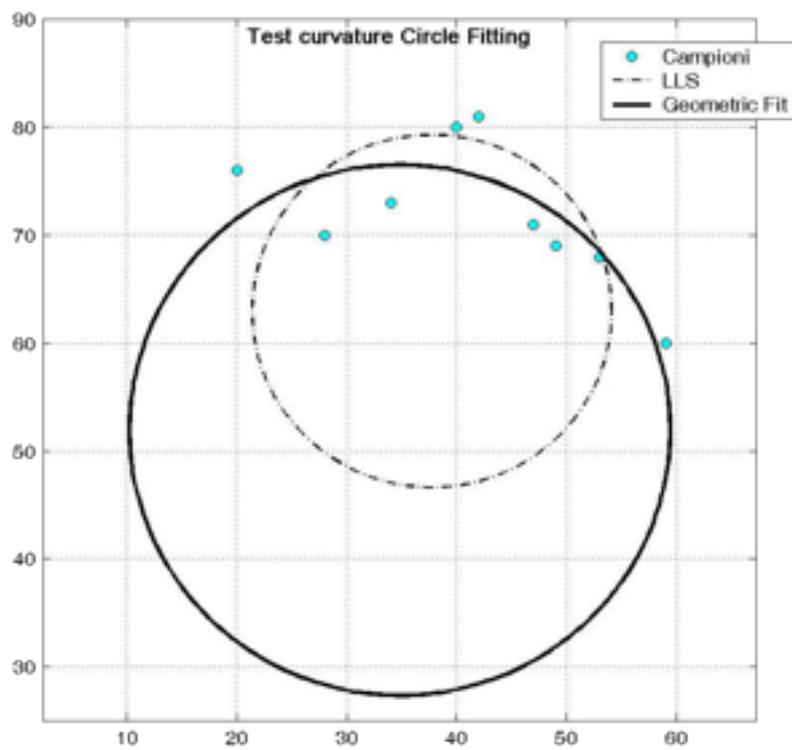


Figura A.10. Metodo LLS e metodo geometrico

A.3 Prodotto di Kronecker

Definizione.(Prodotto di Kronecker) Siano A una matrice $m \times n$ e B una matrice $q \times p$. Il prodotto di Kronecker di A e B è la matrice $mq \times np$ definita da:

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{1n}B \\ \vdots & \vdots \\ a_{m1}B & a_{mn}B \end{bmatrix} \quad (\text{A.25})$$

Proprietà:

1. $(A \otimes B) \otimes C = A \otimes (B \otimes C)$
2. $(A \otimes B)(C \otimes D) = AC \otimes BD$
3. $A \otimes B \neq B \otimes A$
4. $(A \otimes B)^T = A^T \otimes B^T$
5. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$

Definizione.(Vettorizzazione). La vettorizzazione di una matrice A $m \times n$, denotata da $vec(A)$, è il vettore $mn \times 1$ che si ottiene impilando le colonne di A una sotto l'altra.

La connessione tra il prodotto di Kronecker e la vettorizzazione è data dalla relazione:

$$vec(AXB) = (B^T \otimes A) vec(X) \quad (\text{A.26})$$

con A, B, X matrici di dimensioni compatibili.

A.4 Decomposizione ai valori singolari

Teorema: Sia A una matrice $m \times n$. Esiste una matrice D $m \times n$ con elementi diagonali positivi $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$ ed elementi nulli altrove, una matrice U $m \times m$ ortogonale ed una matrice V $n \times n$ ortogonale tali che:

$$U^T AV = D \quad (\text{A.27})$$

Gli elementi diagonali di D prendono il nome di valori singolari.

A.5 Prodotto esterno

Definizione (Prodotto esterno). Il prodotto esterno di due vettori $a, b \in R^3$ si definisce come il vettore:

$$a \times b = \begin{bmatrix} \det \begin{bmatrix} a_2 & a_3 \\ b_2 & b_3 \end{bmatrix} \\ - \det \begin{bmatrix} a_1 & a_3 \\ b_1 & b_3 \end{bmatrix} \\ \det \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \end{bmatrix} \quad (\text{A.28})$$

A.6 Metodo DLT

Tale metodo permette la risoluzione di un sistema lineare omogeneo.

Considerati n punti non complanari, ciascuna corrispondenza tra un punto dell'immagine $m_i = [u_i, v_i, 1]$ ed il punto della scena M_i , fornisce una coppia di equazioni:

$$\begin{cases} p_1^T M_i - u_i p_3^T M_i = 0 \\ p_2^T M_i - v_i p_3^T M_i = 0 \end{cases} \quad (\text{A.29})$$

Che in forma matriciale è:

$$\begin{bmatrix} M_i^T & 0 & -u_i M_i^T \\ 0 & -M_i^T & v_i M_i^T \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = 0_{2 \times 1}. \quad (\text{A.30})$$

Per n punti otteniamo un sistema di $2n$ equazioni lineari omogenee, che possiamo scrivere come:

$$Avec(P^T) = 0 \quad (\text{A.31})$$

dove A è la matrice $2n \times 12$ dei coefficienti e dipende dalle coordinate dei punti, mentre il vettore delle incognite $vec(P^T)$ contiene 12 elementi di P lette per righe. In teoria sei punti, non complanari, sono sufficienti per il calcolo di P , nella pratica sono disponibili molti più punti per compensare gli errori di misura.

Il sistema A.31 è risolto ai minimi quadrati, e la soluzione è l'autovettore associato al minimo autovalore di $A^T A$, che si può calcolare sfruttando la

SVD(appendice A.4 di A.

Notiamo che la derivazione del metodo DLT è rappresentabile sfruttando il prodotto di Kronecker(appendice A.3). Considerate le corrispondenze tra i punti dell'immagine m_i e i punti della scena M_i , si richiede di calcolare una matrice P tale che:

$$m_i = PM_i \quad i = 1..n \quad (\text{A.32})$$

Per eliminare il fattore di scala ricorriamo al prodotto esterno (appendice A.5), riscrivendo come

$$m_i \times PM_i = 0 \quad (\text{A.33})$$

Quindi usando la relazione A.26 si ha:

$$\begin{aligned} m_i \times PM_i = 0 &\Leftrightarrow [m_i]_{\times} PM_i = 0 \Leftrightarrow \\ \text{vec}([m_i]_{\times} PM_i) = 0 &\Leftrightarrow (M_i^T \otimes [m_i]_{\times}) \text{vec}(P) = 0 \end{aligned} \quad (\text{A.34})$$

Abbiamo tre equazioni in 12 incognite, ma solo due di loro sono indipendenti, infatti il rango di $(M_i^T \otimes [m_i]_{\times})$ è due poiché è il prodotto di Kronecker di una matrice di rango uno con una matrice di rango due.

Trasponendo l'equazione A.34 otteniamo:

$$([m_i]_{\times} \otimes M_i^T) \text{vec}(P^T) = 0 \quad (\text{A.35})$$

ed esplicitando la matrice dei coefficienti si ottiene:

$$\begin{bmatrix} 0^T & -M_i^T & v_i M_i^T \\ M_i^T & 0^T & -u_i M_i^T \\ v_i M_i^T & u_i M_i^T & 0^T \end{bmatrix} \text{vec}(P^T) = 0 \quad (\text{A.36})$$

A.7 Trasformata Discreta Coseno

La trasformata DCT permette di sviluppare un segnale, in questo caso un'immagine, mediante una base di soli coseni, a differenza della DFT, che è uno sviluppo su base seni e coseni. Matematicamente la relazione della DCT è:

$$F(u, v) = \sqrt{\frac{2}{M}} \cdot \sqrt{\frac{2}{N}} \cdot \alpha_u \cdot \alpha_v \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \cdot \cos \left[\frac{(2x+1)u\pi}{2M} \right] \cdot \cos \left[\frac{(2y+1)v\pi}{2N} \right] \quad (\text{A.37})$$

dove $f(x, y)$ è il valore dell'immagine, M ed N rappresentano, rispettivamente, il numero di righe e colonne dell'immagine. Inoltre:

$$\alpha_u = \begin{cases} \frac{1}{\sqrt{2}} & \text{se } u = 0 \\ 1 & \text{se } u = 1, 2, \dots, N-1 \end{cases} \quad (\text{A.38})$$

$$\alpha_v = \begin{cases} \frac{1}{\sqrt{2}} & \text{se } v = 0 \\ 1 & \text{se } v = 1, 2, \dots, N-1 \end{cases}$$

La trasformazione inversa IDCT, restituisce l'immagine ricostruita nel dominio spaziale, ed è data dalla relazione:

$$f(x, y) = \sqrt{\frac{2}{M}} \cdot \sqrt{\frac{2}{N}} \cdot \alpha_u \cdot \alpha_v \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v) \cdot \cos \left[\frac{(2x+1)u\pi}{2M} \right] \cdot \cos \left[\frac{(2y+1)v\pi}{2N} \right] \quad (\text{A.39})$$

A.8 Trasformata Discreta Wavelet

La trasformata wavelet è un potentissimo strumento matematico che permette, sfruttando le sue innumerevoli proprietà, di effettuare in modo dettagliato, l'analisi e l'elaborazione di immagini digitali.

Le caratteristiche principali, che contraddistinguono la trasformata wavelet, permettono di:

- analizzare segnali con caratteristiche variabili nel tempo, ad esempio immagini naturali,
- realizzare una rappresentazione multirisoluzione, cioè su diverse scale,
- fornire una loro semplice realizzazione tramite un banco di filtri.

Adesso illustreremo la trasformata wavelet continua, il passaggio alla sua versione discreta, la sua rappresentazione attraverso un banco di filtri ed

il legame con le tecniche di marchiatura digitale. La Trasformata Wavelet Continua (CWT) del segnale $x(t)$ è definita come

$$CWT(x(t)) = CWT(a, b) = \int_{-\infty}^{+\infty} x(t) \psi_b^a(t) dt \quad (\text{A.40})$$

con

$$\psi_b^a(t) = \frac{1}{\sqrt{a}} \left(\frac{t-b}{a} \right) \quad (\text{A.41})$$

detta wavelet madre, dove

$$a \in \mathfrak{R}^+$$

è il parametro di scala e

$$b \in \mathfrak{R}$$

è il parametro di traslazione.

L'analisi tramite CWT opera per successive traslazioni e scalamenti della wavelet, per ogni valore del parametro di scala a il parametro di traslazione b viene fatto variare, consentendo un'analisi dell'intero segnale a risoluzioni diverse.

In particolare:

- • valori grandi del parametro di scala equivalgono ad allungare la wavelet e quindi il suo supporto, lungo l'asse dei tempi, e ad diminuire, contemporaneamente, la frequenza della wavelet; questo permette di ottenere informazioni con margine di precisione più elevato sull'asse delle frequenze rispetto a quello dei tempi,
- • valori piccoli del parametro di scala equivalgono ad comprimere la wavelet e quindi il suo supporto, lungo l'asse dei tempi, e ad aumentare, contemporaneamente, la frequenza della wavelet; questo permette di ottenere informazioni con margine di precisione più elevato sull'asse dei tempi rispetto a quello delle frequenze.

La trasformata wavelet è invertibile, cioè esiste una formula di sintesi che consente di ricostruire il segnale a partire dai coefficienti della trasformata, per segnali $x(t) \in L^2(\mathfrak{R})$ vale la relazione:

$$x(t) = \frac{1}{C_\psi} \int \int_{-\infty}^{+\infty} \frac{1}{a^2} CWT(a, b) \psi_b^a(t) da db \quad (\text{A.42})$$

dove C_ψ è una costante e dipende dalla wavelet scelta.

Poiché nella realtà fisica si lavora con segnali discreti, siamo interessati alla versione discreta della trasformata wavelet, ottenuta forzando il parametro di scale e il parametro di traslazione ad assumere valori discreti:

$$\begin{cases} a = a_0^{-m} & a_0 > 1 & m \in \mathbb{Z} \\ b = nb_0a_0^{-m} & b_0 > 0 & n \in \mathbb{Z} \end{cases} \quad (\text{A.43})$$

da cui sostituendo nell'equazione A.41 e considerando il caso diadico $a_0 = 2$ e $b_0 = 1$ otteniamo

$$\psi_n^j(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \Big|_{a=2^{-j}, b=2^{-j}n} = 2^{j/2} \psi(2^j t - n) \quad (\text{A.44})$$

che rappresenta la wavelet madre discreta, utilizzata nella trasformata wavelet discreta DWT. La trasformata wavelet può essere facilmente generalizzata nel caso bidimensionale, ovvero alle applicazioni che riguardano l'elaborazione delle immagini, compresa la marchiatura digitale.

La trattazione matematica, realizzata per il caso monodimensionale, si estende al caso bidimensionale, si noti inoltre che la funzione padre e la funzione madre sono separabili, quindi è possibile operare prima lungo le righe e poi lungo le colonne del segnale.

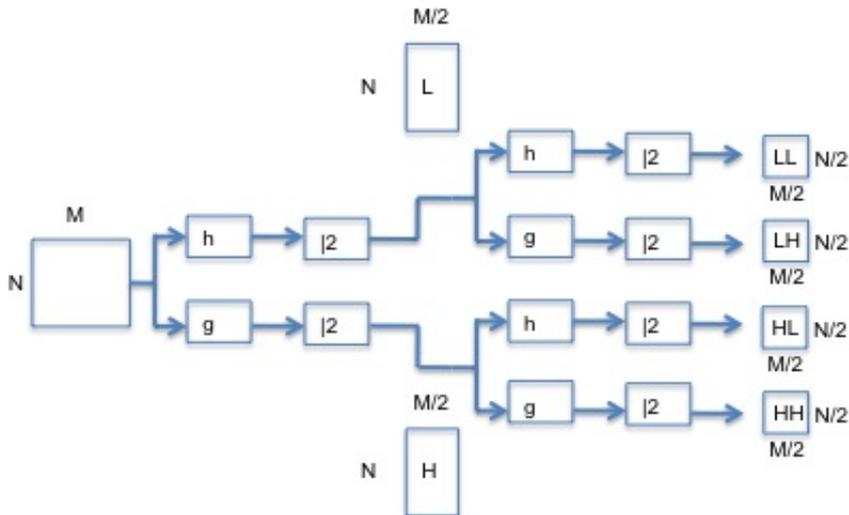


Figura A.11. Schema a blocchi della decomposizione Wavelet

Dal grafico dello schema a blocchi della decomposizione wavelet si nota che a trasformata viene applicata prima lungo le righe, ottenendo sul ramo

superiore i coefficienti dovuti al filtraggio passa-basso e in quello inferiore i coefficienti dovuti al filtraggio passa-alto. Analogamente si esegue la stessa procedura per le colonne. Il risultato sarà dato da una matrice composta da quattro bande:

- banda LL contiene le basse frequenze del segnale, cioè, nel caso di immagini, una versione grossolana dell'immagine stessa,
- banda LH contiene i coefficienti relativi ai dettagli verticali,
- banda HL contiene i coefficienti relativi ai dettagli orizzontali,
- banda HH contiene i coefficienti relativi ai dettagli diagonali.

A.9 Algoritmo di Gauss-Seidel

Il metodo di Gauss-Seidel è un metodo iterativo utilizzato per la risoluzione di un sistema lineare, scritto nella forma matriciale $Ax = b$. La matrice A viene scritta come differenza di due matrici, $A = L - U$, dove L è matrice triangolare inferiore, e U è una matrice triangolare superiore con diagonale nulla. Definendo $A = (a_{ij})_{ij}$ $b = (b_i)_i$, per risolvere il sistema lineare si utilizza quindi una successione $x^{(k)}$ (A.45) che converge verso la soluzione x del sistema lineare.

$$x_i^{(k+1)} = \frac{b_i - \sum_{j < i} a_{ij} x_j^{(k+1)} - \sum_{j > i} a_{ij} x_j^{(k)}}{a_{ii}} \quad (\text{A.45})$$

Indice delle Pubblicazioni

- [Pao01] Luca Costantini ; Paolo Sità ; Licia Capodiferro ; Alessandro Neri; Laguerre Gauss analysis for image retrieval based on color texture. Proc. SPIE 7535, Wavelet Applications in Industrial Processing VII, 75350G (February 04, 2010);
- [Pao02] VI CONGRESO INTERNACIONAL de Musealización de Yacimientos Arqueológicos y Patrimonio, Toledo, Spain November 22-25, 2010 A. Ancona, A. Contino, L. D'Alessandro, F. Riccio, R. Sebastiani, A. Capodiferro, G. Ciccarello, E. Di Carlo, P. Quaranta, L. Capodiferro, C. Delogu, E. Pallotti, P. Sità Il museo diffuso del rione Testaccio e il colle Aventino a Roma: la comunicazione dei beni culturali tra realtà e multimedialità
- [Pao03] L. Costantini ; P. Sità ; M. Carli ; A. Neri; Features extraction based on Fisher's information. Proc. SPIE 7870, Image Processing: Algorithms and Systems IX, 787016 (February 03, 2011); doi:10.1117/12.876581
- [Pao04] E. Pallotti; L. Capodiferro; F. Mangiatordi; P. Sità, *Smooth image inpainting by least square oriented edge prediction*, SPIE Proceedings Vol. 8295 Image Processing: Algorithms and Systems X, DOI:10.1117/12.908297, March 2012.
- [Pao05] Costantini, L.; Mangiatordi, F.; Pallotti, E.; Sita, P., CHIP — Cultural heritage image processing tool, Communications Control and Signal Processing (ISCCSP), 2012 5th International Symposium on , vol., no., pp.1,6, 2-4 May 2012 doi: 10.1109/ISCCSP.2012.6217820

Bibliografia

- [1] C. Wu, “Towards linear-time incremental structure from motion,” in *3D Vision, 2013 International Conference on*, pp. 127–134, June 2013.
- [2] Y. Furukawa and J. Ponce, “Accurate, dense, and robust multi-view stereopsis,” in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pp. 1–8, June 2007.
- [3] M. Jancosek and T. Pajdla, “Multi-view reconstruction preserving weakly-supported surfaces,” in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 3121–3128, June 2011.
- [4] M. A. Fischler and R. C. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Commun. ACM*, vol. 24, pp. 381–395, June 1981.
- [5] D. Marquardt, “An algorithm for least-squares estimation of nonlinear parameters,” *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.
- [6] G.Emilie-Male, *The Restorer’s Handbook of Easel Painting*. Van Nostrand Reinhold, 1976.
- [7] R. Arnheim, “Untersuchungen zur lehre von der gestalt,” *Psychologische Forschung*, vol. 11, no. 1, pp. 2–119, 1928.
- [8] S. Rane, J. Remus, and G. Sapiro, “Wavelet-domain reconstruction of lost blocks in wireless image transmission and packet-switched networks,” in *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 1, pp. I–309–I–312 vol.1, 2002.

- [9] D. Zhang and Z. Wang, "Image information restoration based on long-range correlation," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 12, pp. 331–341, May 2002.
- [10] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '00*, (New York, NY, USA), pp. 417–424, ACM Press/Addison-Wesley Publishing Co., 2000.
- [11] W. Gander, G. Golub, and R. Strebler, "Least-squares fitting of circles and ellipses," *BIT Numerical Mathematics*, vol. 34, no. 4, pp. 558–578, 1994.
- [12] R. Frankot and R. Chellappa, "A method for enforcing integrability in shape from shading algorithms," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 10, pp. 439–451, Jul 1988.
- [13] G. Jacovitti and A. Neri, "Multiresolution circular harmonic decomposition," *Signal Processing, IEEE Transactions on*, vol. 48, pp. 3242–3247, Nov 2000.
- [14] L. Capdiferro, V. Casieri, A. Laurenti, and G. Jacovitti, "Multiple feature based multiscale image enhancement," in *Digital Signal Processing, 2002. DSP 2002. 2002 14th International Conference on*, vol. 2, pp. 931–934 vol.2, 2002.
- [15] L. Capodiferro, E. Di Claudio, F. Iacolucci, A. Laurenti, and G. Jacovitti, "Two-channel technique for high dynamic range image visualization," in *Information Visualisation, 2004. IV 2004. Proceedings. Eighth International Conference on*, pp. 269–273, July 2004.
- [16] D. V. Sorokin, M. M. Mizotin, and A. S. Krylov, "Gauss-laguerre keypoints extraction using fast hermite projection method," in *Proceedings of the 8th International Conference on Image Analysis and Recognition - Volume Part I, ICIAR'11*, (Berlin, Heidelberg), pp. 284–293, Springer-Verlag, 2011.
- [17] A. Shortt, T. Naughton, and B. Javidi, "Histogram approaches for lossy compression of digital holograms of three-dimensional objects," *Image Processing, IEEE Transactions on*, vol. 16, pp. 1548–1556, June 2007.

-
- [18] S. Dolev and S. Frenkel, "Multiplication free holographic coding," in *Electrical and Electronics Engineers in Israel (IEEEI), 2010 IEEE 26th Convention of*, pp. 000146–000150, Nov 2010.
- [19] M. Kroll, L. Muhlfeld, and D. Block, "Stereoscopic digital holography," *Plasma Science, IEEE Transactions on*, vol. 38, pp. 897–900, April 2010.
- [20] S. Voloshynovskiy, S. Pereira, T. Pun, J. Eggers, and J. Su, "Attacks on digital watermarks: classification, estimation based attacks, and benchmarks," *Communications Magazine, IEEE*, vol. 39, pp. 118–126, Aug 2001.
- [21] Z. Zhang, "A flexible new technique for camera calibration," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, pp. 1330–1334, Nov 2000.
- [22] Y.-Y. Yeh and L. D. Silverstein, "Limits of fusion and depth judgment in stereoscopic color displays," *Hum. Factors*, vol. 32, pp. 45–60, Jan. 1990.
- [23] H. Yamanoue, M. Okui, and I. Yuyama, "A study on the relationship between shooting conditions and cardboard effect of stereoscopic images," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 10, pp. 411–416, Apr 2000.
- [24] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 14, pp. 965–980, Oct 1992.
- [25] R. Tsai, "The robotics review 1," ch. Synopsis of Recent Progress on Camera Calibration for 3D Machine Vision, pp. 147–159, Cambridge, MA, USA: MIT Press, 1989.
- [26] P. H. S. Torr and D. W. Murray, "The development and comparison of robust methods for estimating the fundamental matrix," *Int. J. Comput. Vision*, vol. 24, pp. 271–300, Sept. 1997.
- [27] T. Huang and A. Netravali, "Motion and structure from feature correspondences: a review," *Proceedings of the IEEE*, vol. 82, pp. 252–268, Feb 1994.

- [28] S. Sull and N. Ahuja, “Estimation of motion and structure of planar surfaces from a sequence of monocular images,” in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on*, pp. 732–733, Jun 1991.
- [29] S. J. Maybank and O. D. Faugeras, “A theory of self-calibration of a moving camera,” *Int. J. Comput. Vision*, vol. 8, pp. 123–151, Aug. 1992.
- [30] “Computer vision, graphics, and image processing.”
- [31] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vision*, vol. 60, pp. 91–110, Nov. 2004.
- [32] L. Agapito, E. Hayman, and I. Reid, “Self-calibration of rotating and zooming cameras,” *Int. J. Comput. Vision*, vol. 45, pp. 107–127, Nov. 2001.
- [33] S. Wang and Y. Wu, “The studies and explores visualization in the information of cultural heritage protection,” in *Computer-Aided Industrial Design Conceptual Design (CAIDCD), 2010 IEEE 11th International Conference on*, vol. 2, pp. 980–983, Nov 2010.
- [34] C. Ozmen and S. Balcisoy, “Bi-manual interactive tools for cultural heritage researchers,” in *Cyberworlds, 2008 International Conference on*, pp. 439–446, Sept 2008.
- [35] E. Berndt and J. Carlos, “Cultural heritage in the mature era of computer graphics,” *Computer Graphics and Applications, IEEE*, vol. 20, pp. 36–37, Jan 2000.
- [36] B. Flynn, “Augmented visualisation: Designing experience for an interpretative cultural heritage,” in *Information Visualisation, 2008. IV '08. 12th International Conference*, pp. 447–452, July 2008.
- [37] G. Cimadomo, “Documentation and dissemination of cultural heritage: Current solutions and considerations about its digital implementation,” in *Digital Heritage International Congress (DigitalHeritage), 2013*, vol. 1, pp. 555–562, Oct 2013.

- [38] A. Manferdini, “The use of low-cost technologies for the promotion of cultural heritage sites: The case study of veleia,” in *Virtual Systems and Multimedia (VSMM), 2012 18th International Conference on*, pp. 484–490, Sept 2012.
- [39] Y. Zhang, Q. Ji, and W. Zhang, “Multi-view autostereoscopic 3d display,” in *Optics Photonics and Energy Engineering (OPEE), 2010 International Conference on*, vol. 1, pp. 58–61, May 2010.
- [40] A. Manferdini and M. Russo, “Multi-scalar 3d digitization of cultural heritage using a low-cost integrated approach,” in *Digital Heritage International Congress (DigitalHeritage), 2013*, vol. 1, pp. 153–160, Oct 2013.
- [41] S. Weibel, M. Olbrich, T. Franke, and J. Keil, “Immersive experience of current and ancient reconstructed cultural attractions,” in *Digital Heritage International Congress (DigitalHeritage), 2013*, vol. 1, pp. 395–398, Oct 2013.
- [42] M. Andreetto, N. Brusco, and G. Cortelazzo, “Automatic 3d modeling of textured cultural heritage objects,” *Image Processing, IEEE Transactions on*, vol. 13, pp. 354–369, March 2004.
- [43] F. Mazzenga, A. Simonetta, R. Giuliano, and M. Vari, “Applications of smart tagged rfid tapes for localization services in historical and cultural heritage environments,” in *Enabling Technologies: Infrastructures for Collaborative Enterprises (WETICE), 2010 19th IEEE International Workshop on*, pp. 186–191, June 2010.
- [44] M. Brown and D. Lowe, “Recognising panoramas,” in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pp. 1218–1225 vol.2, Oct 2003.
- [45] C. Harris and M. Stephens, “A combined corner and edge detector,” in *In Proc. of Fourth Alvey Vision Conference*, pp. 147–151, 1988.
- [46] H. Hse and A. Newton, “Sketched symbol recognition using zernike moments,” in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 1, pp. 367–370 Vol.1, Aug 2004.

- [47] A. Neri and G. Jacovitti, "Maximum likelihood localization of 2-d patterns in the gauss-laguerre transform domain: theoretic framework and preliminary results," *Image Processing, IEEE Transactions on*, vol. 13, pp. 72–86, Jan 2004.
- [48] C. Woodley, S. Marshall, S. Taylor, and S. Fagan, "Technologies, indigenous cultural heritage and community capacity building," in *Global Humanitarian Technology Conference (GHTC), 2013 IEEE*, pp. 406–410, Oct 2013.
- [49] Y. Zhao, P. Campisi, and D. Kundur, "Dual domain watermarking for authentication and compression of cultural heritage images," *Image Processing, IEEE Transactions on*, vol. 13, pp. 430–448, March 2004.
- [50] V. Cappellini, F. Bartolini, R. Caldelli, A. D. Rosa, A. Piva, and M. Barni, "Robust frame-based watermarking for digital video," in *Proceedings of the 12th International Workshop on Database and Expert Systems Applications, DEXA '01*, (Washington, DC, USA), pp. 825–829, IEEE Computer Society, 2001.
- [51] A. Piva, M. Barni, F. Bartolini, and V. Cappellini, "Dct-based watermark recovering without resorting to the uncorrupted original image," in *Proceedings of the 1997 International Conference on Image Processing (ICIP '97) 3-Volume Set-Volume 1 - Volume 1*, ICIP '97, (Washington, DC, USA), pp. 520–, IEEE Computer Society, 1997.
- [52] M. Barni, F. Bartolini, and V. Cappellini, "Image processing for virtual restoration of artworks," *IEEE MultiMedia*, vol. 7, pp. 34–37, Apr. 2000.
- [53] V. Cappellini, G. Damiani, and A. Piva, "Protection of cultural heritage contents on the web," in *Proceedings of the Database and Expert Systems Applications, 15th International Workshop, DEXA '04*, (Washington, DC, USA), pp. 57–61, IEEE Computer Society, 2004.
- [54] F. Battisti, M. Cancellaro, G. Boato, M. Carli, and A. Neri, "Joint watermarking and encryption of color images in the fibonacci-haar domain," *EURASIP J. Adv. Signal Process*, vol. 2009, pp. 43:3–43:3, Jan. 2009.

- [55] M. Cancellaro, F. Battisti, M. Carli, G. Boato, F. G. De Natale, and A. Neri, "A commutative digital image watermarking and encryption method in the tree structured haar transform domain," *Signal Processing: Image Communication*, vol. 26, no. 1, pp. 1–12, 2011.
- [56] D. Nister, "An efficient solution to the five-point relative pose problem," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 2, pp. II–195–202 vol.2, June 2003.
- [57] A. Fusiello, V. Roberto, and E. Trucco, "Efficient stereo with multiple windowing," in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pp. 858–863, Jun 1997.
- [58] R. Gherardi, M. Farenzena, and A. Fusiello, "Improving the efficiency of hierarchical structure-and-motion," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 1594–1600, June 2010.
- [59] P. Besl and N. D. McKay, "A method for registration of 3-d shapes," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 14, pp. 239–256, Feb 1992.
- [60] R. Hartley, "In defense of the eight-point algorithm," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, pp. 580–593, Jun 1997.
- [61] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, pp. 1254–1259, Nov 1998.
- [62] M. I. A. Lourakis, M. I. A. Lourakis, A. A. Argyros, and A. A. Argyros, "The design and implementation of a generic sparse bundle adjustment software package based on the levenberg-marquardt algorithm," tech. rep., 2004.
- [63] N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: Exploring photo collections in 3d," in *ACM TRANSACTIONS ON GRAPHICS*, pp. 835–846, Press, 2006.

- [64] H. C. Longuet-Higgins, “Readings in computer vision: Issues, problems, principles, and paradigms,” ch. A Computer Algorithm for Reconstructing a Scene from Two Projections, pp. 61–62, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1987.
- [65] R. I. Hartley, “Self-calibration of stationary cameras,” *International Journal of Computer Vision*, vol. 22, pp. 5–23, 1997.
- [66] M. Naaman, Y. Song, A. Paepcke, and H. Garcia-Molina, “Automatic organization for digital photographs with geographic coordinates,” in *Digital Libraries, 2004. Proceedings of the 2004 Joint ACM/IEEE Conference on*, pp. 53–62, June 2004.
- [67] A. Roman, G. Garg, and M. Levoy, “Interactive design of multi-perspective images for visualizing urban landscapes,” in *Proceedings of the Conference on Visualization '04, VIS '04*, (Washington, DC, USA), pp. 537–544, IEEE Computer Society, 2004.
- [68] N. Snavely, S. M. Seitz, and R. Szeliski, “Modeling the world from internet photo collections,” *Int. J. Comput. Vision*, vol. 80, pp. 189–210, Nov. 2008.
- [69] Y. Furukawa and J. Ponce, “Accurate, dense, and robust multiview stereopsis,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, pp. 1362–1376, Aug 2010.
- [70] R. Fattal, D. Lischinski, and M. Werman, “Gradient domain high dynamic range compression,” *ACM Trans. Graph.*, vol. 21, pp. 249–256, July 2002.