

Ricostruzione Tridimensionale di Volti Umani con lo Stereo Fotometrico

Matteo Fois

21 giugno 2016

Indice

Introduzione	2
1 Motivazioni ed Obiettivi	3
2 Ricerca Correlata	4
2.1 Stereo Binoculare e Multiview	4
2.2 Scannerizzazione Laser	5
2.3 Shape from Shading	6
2.4 Modelli di Formazione d'Immagini	6
2.4.1 Il Modello Lambertiano	7
2.4.2 Il Modello di Phong	8
2.5 Mappe di Riflettanza	8
3 Lo Stereo Fotometrico	11
3.1 Recupero di Normali ed Albedo	12
3.2 Filtraggio di Ombre Proiettate e Riflessioni Speculari	13
3.2.1 Filtro d'Intensità	14
3.2.2 Filtro Fotometrico	15
4 Segmentazione dei Volti	17
4.1 Costruzione delle Griglie	18
4.2 Vantaggi Rispetto alle Griglie Tradizionali	19
5 Integrazione delle Normali	20
5.1 Equazione Eikonale e Schemi Semi-Lagrangiani	21
5.2 Implementazione Fast-Marching	23
5.3 Strategia di Penalizzazione	25
5.3.1 Stima del Parametro di Penalizzazione	26
6 Risultati	28
6.1 Banco di Prova: Il Database di Yale	28
6.2 Alcune Ricostruzioni dei Volti di Yale	30
6.3 Effetti del Filtraggio delle Immagini	31
6.4 Ricostruzioni da Singole Immagini	34
7 Conclusioni e Prospettive Future	36

Introduzione

Si presenta un nuovo approccio al problema di ricostruzione tridimensionale di volti umani da immagini multiple, basato sulla tecnica dello Stereo Fotometrico. I volti impressi nelle immagini sono individuati e segmentati automaticamente da una libreria di riconoscimento facciale di recente concezione, che traccia la posizione di alcuni punti chiave nelle foto di un volto. Con questi punti viene costruita una griglia non strutturata su cui si applica lo Stereo Fotometrico per ricavare il colore e le normali alla superficie dei volti. Le normali ottenute vengono poi integrate con uno schema Semi-Lagrangiano Fast-Marching per recuperare i valori di profondità delle superfici.

Per superare alcuni limiti dello Stereo Fotometrico, nello specifico il trattamento di ombre proiettate e specularità nelle immagini, vengono proposti dei filtri in grado di selezionare per ogni punto della griglia le immagini migliori da utilizzare.

Il metodo e i filtri sono messi alla prova ricostruendo i volti del database dell'università di Yale. Si riportano infine alcuni esempi di ricostruzione di volti da singole immagini, ottenuti combinando le normali di un volto template con la griglia e il colore delle immagini.

Capitolo 1

Motivazioni ed Obiettivi

Questo lavoro è nato con l'obiettivo di sviluppare un applicativo in grado di costruire un modello tridimensionale del volto di un utente con immagini provenienti da una fotocamera o da una webcam. Ottenere delle ricostruzioni fedeli di un volto utilizzando delle immagini è un risultato importante, che può portare allo sviluppo di applicazioni utili in campi come la grafica, per l'animazione e la produzione di nuovi contenuti, e la medicina, per l'analisi di nei potenzialmente maligni o di parti di volto che un chirurgo plastico deve operare.

Nelle intenzioni originali, il modello ottenuto doveva inoltre poter essere animato, trasmettendogli tramite una webcam movimento ed espressione. Gli algoritmi di riconoscimento facciale di ultima generazione consentono infatti lo sviluppo di applicazioni di motion capture che non richiedono l'uso di sensori fisici, aprendo la strada ad una miriade di nuove tecniche nell'ambito dell'animazione tridimensionale.

Data la flessibilità e le potenzialità offerte da questi nuovi strumenti, l'obiettivo principale che questa ricerca si è posta è diventato poi quello di sperimentarne la combinazione con alcune tecniche tradizionali di ricostruzione tridimensionale da immagini, partendo nello specifico dalla tecnica dello Stereo Fotometrico.

Trattandosi di un approccio relativamente nuovo, sono emerse nel corso del lavoro diverse sfide di implementazione, legate in particolare alla struttura della griglia utilizzata per segmentare i volti, che hanno richiesto lo sviluppo di un metodo numerico di integrazione capace di operare su griglie non strutturate.

Capitolo 2

Ricerca Correlata

La ricostruzione tridimensionale di oggetti da immagini è fra i campi più attivi nella ricerca in grafica computerizzata, ed ha determinato, nel corso degli ultimi decenni, la nascita di diverse tecniche e strumenti per affrontare il problema. In questo capitolo vengono descritti brevemente i tre filoni di ricerca principali (con l'eccezione della tecnica dello Stereo Fotometrico, che viene trattata in un capitolo a parte): lo Stereo Binoculare (o Multiview), la scannerizzazione laser e lo Shape from Shading. Per ognuno di questi metodi vengono riportati inoltre alcuni dei risultati più recenti legati alla ricostruzione di volti umani, con l'obiettivo di fornire al lettore alcune informazioni sullo stato dell'arte della ricerca e poter fare dei confronti con i risultati ottenuti in questo lavoro. Vengono infine proposti con alcuni accenni di teoria di modelli di riflessione della luce e formazione di immagini.

2.1 Stereo Binoculare e Multiview

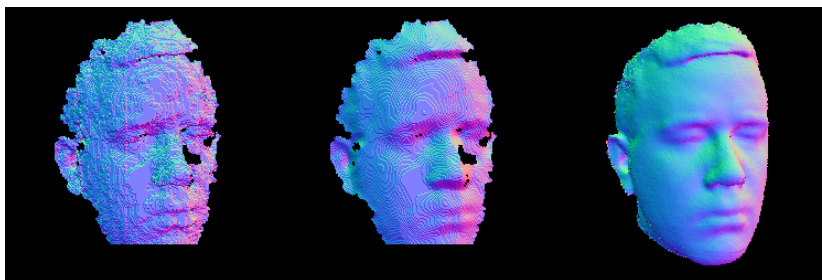


Figura 2.1: Tre gradi di ricostruzione di un volto ottenuti con la doppia fotocamera di un Kinect usando la tecnica dello Stereo Multiview.

Lo Stereo Binoculare è una tecnica che consente di determinare i valori di profondità dei punti di una superficie utilizzando due immagini di questa prese da due differenti punti di vista, come mostrato in figura.

Mimando il funzionamento della visione binoculare umana, la tecnica agisce individuando inizialmente le corrispondenze tra i punti delle due immagini. Una volta individuata una corrispondenza, per esempio tra il pixel (x_1, y_1) nella pri-

ma immagine e il pixel (x_2, y_2) nella seconda, si fa uso della distanza tra i due punti di vista per triangolare la profondità del punto della superficie impresso nei due pixel. In un lavoro di A. Chambolle, riportato in [1], viene dimostrato che il problema con due immagini è ben posto ed ammette soluzione unica. La tecnica può impiegare più di due punti vista (prendendo il nome più generale di Stereo Multiview), utilizzando le immagini aggiuntive per testare la bontà della ricostruzione e raffinarne i risultati.

Per quanto riguarda la ricostruzione tridimensionale di volti da immagini, lo Stereo Multiview è uno degli strumenti su cui si la ricerca si è più concentrata negli ultimi anni, in particolare dall'introduzione nel mercato casalingo di fotocamere stereo come quella del Microsoft Kinect o del Nintendo 3DS, che con i loro due obiettivi sono in grado rendere possibile l'applicazione dello Stereo Multiview anche al di fuori dei ambienti specializzati.

2.2 Scannerizzazione Laser

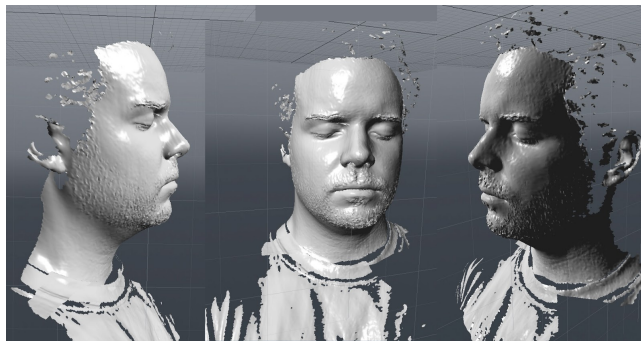


Figura 2.2: Nuvola di punti di un volto ottenuta con uno scanner laser.

Lo scanner laser è uno strumento elettronico che fornisce le coordinate spaziali di una nuvola di punti appartenenti alla scena sottoposta al rilievo. Il rilievo avviene utilizzando dei *distanziometri*, oggetti in grado di emettere raggi laser e misurare la distanza dal punto di intersezione tra il raggio emesso ed una superficie opaca.

Il risultato finale è simile ad una fotografia, con la differenza che invece del colore dei punti degli oggetti ripresi, lo scanner ne registra la distanza.

La nuvola di punti ottenuta può essere poi usata per estrapolare la forma degli oggetti rilevati.

Come una fotocamera, lo scanner è in grado di registrare informazioni solo sui punti che appaiono nel suo campo di vista. Per ottenere una ricostruzione a tutto tondo di un oggetto si possono combinare con un processo di allineamento più rilievi, presi da direzioni differenti, in maniera analoga alla tecnica dello Stereo Multiview.



Figura 2.3: Ricostruzione del volto di Lena con lo Shape from Shading.

2.3 Shape from Shading

Lo Shape from Shading (dall'inglese "forma da ombreggiatura") cerca di risolvere il problema di ricostruire tridimensionalmente un oggetto da una singola immagine, sfruttando la relazione tra l'intensità di colore registrata nell'immagine e l'orientazione delle normali alla superficie, data da un modello matematico di riflettanza della luce.

E' noto che lo Shape from Shading è un problema mal posto, ovvero è impossibile determinare in maniera univoca le normali alla superficie di un oggetto da una singola immagine senza aggiungere ulteriori informazioni, poichè esistono infinite superfici che producono la stessa immagine quando illuminate dalla stessa sorgente luminosa. Il problema richiede inoltre la conoscenza a priori del colore della superficie da ricostruire (che in molti studi viene assunto costante) e delle condizioni al bordo, ovvero dei valori di profondità lungo il contorno della superficie. Queste informazioni sono raramente disponibili, e limitano l'applicazione dello Shape from Shading a situazioni molto restrittive.

Per quanto riguarda l'applicazione dello Shape from Shading alla ricostruzione di volti umani, i risultati ottenuti dipendono sensibilmente dalla conoscenza a priori del colore dei volti, e l'assunzione di un colore costante determina risultati completamente incorretti, come mostrato in figura 2.3.

Una ricerca interessante, riportata in [9], propone di risolvere le ambiguità dello Shape from Shading utilizzando un volto template, ottenuto ad esempio tramite scannerizzazione, che fornisce una stima iniziale sul colore del volto da ricostruire e sull'orientazione delle normali alla superficie. Il volto template viene poi deformato sulla base delle informazioni sull'ombreggiatura contenute in una singola immagine.

2.4 Modelli di Formazione d'Immagini

L'aspetto della superficie di un oggetto catturata in un'immagine è il risultato di una combinazione della forma dell'oggetto, delle condizioni di illuminazione della scena e delle sue caratteristiche di riflettanza, ovvero della proporzione di luce incidente che la superficie è in grado di riflettere.

Per caratterizzare le intensità luminose registrate in un'immagine rispetto alla

luce, al punto di vista dell'osservatore e alle proprietà ottiche dei materiali degli oggetti raffigurati, sono stati sviluppati diversi modelli matematici di riflettanza. Tali modelli offrono una descrizione locale dei meccanismi di riflessione della luce che può rivelarsi utile come base per la ricostruzione tridimensionale delle superfici impresse in un'immagine.

La frazione di luce I riflessa dalla superficie di un oggetto in una data direzione dipende in generale dalle proprietà ottiche e dalle microstrutture del materiale di cui è composta. Tuttavia per molte superfici, questa frazione dipende solo dalla loro orientazione locale. Le caratteristiche di riflettanza di tali superfici possono essere rappresentate punto per punto da una funzione della normale esterna alla superficie nel punto \vec{n} , della direzione della luce incidente \vec{l} e del punto di vista \vec{v} dell'osservatore

$$I = \phi(\vec{n}, \vec{l}, \vec{v}),$$

Questa funzione determina il rapporto tra luminosità incidente e luminosità riflessa nella direzione dell'osservatore, misurato per unità di area superficiale.

2.4.1 Il Modello Lambertiano

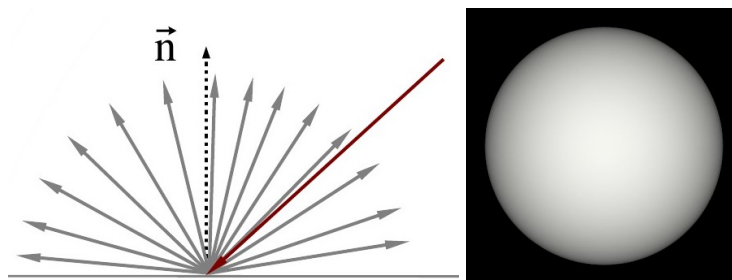


Figura 2.4: A sinistra, rappresentazione grafica del modello di diffusione lambertiana: la luce incidente (freccia rossa) viene diffusa uniformemente in tutte le direzioni, con intensità proporzionale al coseno dell'angolo che forma con la normale esterna. A destra, una superficie sferica lambertiana.

Uno dei modelli più semplici di riflettanza è il modello lambertiano, proposto dal matematico Johann Heinrich Lambert intorno alla metà del 1700. Secondo questo modello, la frazione di luce incidente riflessa da una superficie varia come il coseno dell'angolo tra la normale alla superficie e la direzione dei raggi di luce, e questa quantità è la stessa per tutte le direzioni, ovvero quando un elemento d'area della superficie è visto da qualsiasi angolo si osserva sempre la stessa luminosità apparente. Il modello ha dunque la forma

$$I = \rho \vec{n} \cdot \vec{l}$$

dove la costante ρ , che rappresenta il colore o l'*albedo* della superficie, è un fattore di riflettanza tipico del materiale di cui è composta. Una superficie che obbedisce a questo modello di riflettanza si dice lambertiana.

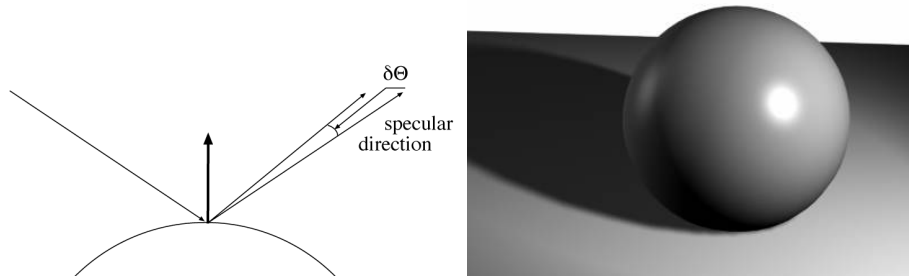


Figura 2.5: A sinistra, rappresentazione grafica del modello di diffusione speculare: la luce incidente viene riflessa in un piccolo intorno della direzione speculare. A destra, una superficie sferica speculare

2.4.2 Il Modello di Phong

Il modello di Phong nasce per modellare dei fenomeni non lambertiani di riflessione della luce, nello specifico le cosiddette *riflessioni speculari*, riflessioni che determinano un'intensità di luce concentrata intorno ad una particolare direzione, la direzione speculare, che è simmetrica a quella della luce incidente rispetto all'asse su cui giace la normale alla superficie, come mostrato in figura 2.5. Una superficie che esibisce questo tipo di riflessioni si dice essere speculare, o anche un riflettore.

Per un riflettore ideale, come ad esempio uno specchio, le riflessioni speculari sono visibili da un osservatore solo se le direzioni speculari coincidono con quella del punto di vista. Per oggetti reali, queste riflessioni sono visibili quando la direzione del punto di vista è contenuta in un cono di direzioni di una certa ampiezza intorno alla direzione speculare. La loro intensità risulta massima quando l'angolo $\delta\Theta$ tra la direzione del punto di vista e la direzione speculare è uguale a zero, e decresce progressivamente all'aumentare di $\delta\Theta$.

Il modello proposto dunque da Phong per modellare le riflessioni speculari ha la forma

$$I = \cos^n(\delta\Theta),$$

dove l'esponente n ha il ruolo di regolare l'ampiezza del cono di direzioni per le quali è visibile la riflessione.

2.5 Mappe di Riflettanza

Si consideri l'immagine di un oggetto posto idealmente su un piano (x, y) , presa con proiezione ortografica (per evitare distorsioni prospettiche) da un certo punto di vista con direzione \vec{v} perpendicolare al piano.

L'immagine, che si può pensare per semplicità in bianco e nero, viene spesso rappresentata in maniera discreta da una matrice $(I_{i,j})$, dove i e j sono le coordinate orizzontali e verticali di un pixel e il valore $I_{i,j} \in \{0, \dots, 255\}$ denota l'intensità, o il tono di grigio, del pixel, ma è possibile e talvolta più comodo darne un'interpretazione nel continuo introducendo le coordinate immagine (u, v) e pensandola come una funzione $I(u, v)$ a valori nell'intervallo $[0, 1]$. Si può

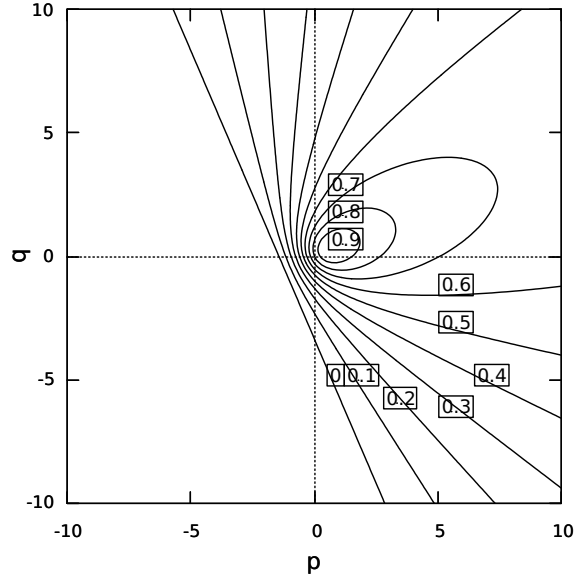


Figura 2.6: Grafico delle curve di livello di una mappa di riflettanza ottenuta con il modello di diffusione lambertiana.

assumere una scala appropriata del piano dell'immagine, tale che le coordinate (x, y, z) dei punti della superficie dell'oggetto vengono mappate sulle coordinate (u, v) dell'immagine in modo che si abbia $u = x$, $v = y$. Se la superficie dell'oggetto è rappresentabile esplicitamente da un'equazione del tipo

$$z = f(x, y)$$

allora la normale esterna alla superficie è data in ogni punto da

$$\vec{n} = \frac{(-p, -q, -1)}{\sqrt{1 + p^2 + q^2}}$$

dove

$$p = \frac{\partial f}{\partial x}, \quad q = \frac{\partial f}{\partial y}$$

Si supponga dunque che l'oggetto impresso nell'immagine sia illuminato da un'unica sorgente luminosa, posta ad una distanza abbastanza grande da poter considerare i raggi di luce incidenti sulla superficie tutti provenienti dalla stessa direzione

$$\vec{l} = (l_1, l_2, l_3).$$

Avendo fissato il punto di vista dell'osservatore, la funzione $\phi(\vec{n}, \vec{l}, \vec{v})$ che descrive in ogni punto le caratteristiche di riflettanza della superficie dipende solo dalla sua orientazione locale, data dalla normale esterna \vec{n} . In questo caso ϕ è una funzione dello spazio gradiente (p, q) e viene definita una *Mappa di Riflettanza*,

spesso indicata con la notazione $R(p, q)$. Per ogni punto (x, y) dell'immagine si ha dunque

$$I(x, y) = R(p, q)$$

Nella pratica, una mappa di riflettanza può essere determinata empiricamente analizzando le immagini, derivata da modelli fenomenologici della riflettanza della superficie o da modelli analitici delle microstrutture del materiale di cui la superficie è composta. A titolo di esempio, la mappa di riflettanza corrispondente al modello lambertiano è data da

$$R(p, q) = \rho \frac{(-pl_1 - ql_2 + l_3)}{\sqrt{1 + p^2 + q^2}}.$$

Capitolo 3

Lo Stereo Fotometrico

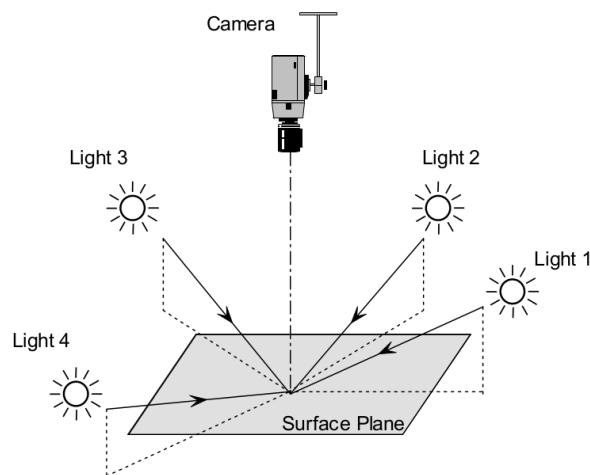


Figura 3.1: Schema grafico della tecnica dello stereo fotometrico: diverse immagini di una superficie vengono scattate da uno stesso punto di vista facendo variare la posizione della sorgente luminosa.

La tecnica dello Stereo Fotometrico, introdotta da un articolo di Robert Woodham in [15], permette di calcolare il colore e l'orientazione locale di una superficie facendo uso di più immagini di questa prese dallo stesso punto di vista, facendo variare la direzione l della luce incidente tra un'immagine e l'altra. Questa tecnica è stata utilizzata nel presente lavoro per ricostruire volti umani da una serie di immagini, per cui si è scelto di dedicare questo capitolo ad una descrizione più accurata della teoria matematica su cui la tecnica è basata. Supponendo che la superficie impressa nelle immagini sia lambertiana e che l'albedo ρ e le direzioni della sorgente luminosa siano note, sono sufficienti due sole immagini per determinare l'orientazione della superficie in ogni punto. Alle due immagini I_1, I_2 corrispondono due diverse mappe di riflettanza $R_1(p, q), R_2(p, q)$, da cui si ottiene un sistema di due equazioni indipendenti

$$I_1(x, y) = R_1(p, q)$$

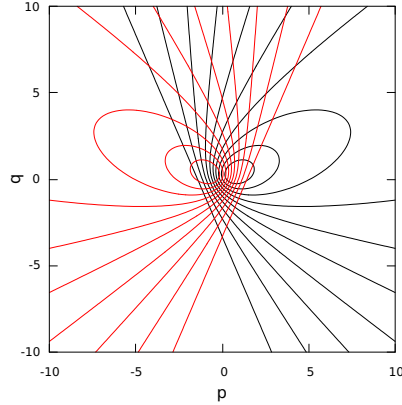


Figura 3.2: Intersezioni tra le curve di livello di due mappe di riflettanza.

$$I_2(x, y) = R_2(p, q)$$

che può essere risolto rapidamente con il metodo di Cramer.

Il meccanismo di base della tecnica può essere rappresentato utilizzando i grafici delle due mappe di riflettanza, come mostrato in figura 3.2. Le coppie (p, q) che risolvono il sistema in ogni punto della superficie non sono altro che le intersezioni delle curve di livello di $R_1(p, q)$ ed $R_2(p, q)$ corrispondenti ai livelli di intensità delle due immagini.

Se l'angolo di fase g tra la direzione della luce e il punto di vista è lo stesso per entrambe le immagini (ovvero se la direzione dell'illuminazione viene fatta ruotare attorno all'asse del punto di vista) allora le due mappe di riflettanza sono rotazioni di una stessa mappa, e una volta nota la prima, il calcolo dell'altra è immediato. Questa semplificazione risulta molto utile per applicazioni della tecnica in ambito industriale, dove ad esempio si vuole determinare l'orientazione della superficie di un oggetto che deve soddisfare degli standard di produzione. Avendo fissati le direzioni della sorgente luminosa e l'albedo dell'oggetto, le mappe di riflettanza possono essere computate una volta per tutte, e le loro intersezioni vanno a costituire una tabella di lookup del tipo $(I_1, I_2) = (p, q)$, che può essere facilmente consultata dopo aver preso le immagini per controllare che l'orientazione sia quella giusta.

3.1 Recupero di Normali ed Albedo

In generale l'albedo ρ della superficie da ricostruire non è noto a priori, e potrebbe differire punto per punto a seconda del materiale e della sua texture, per cui è opportuno considerarlo come una funzione della posizione $\rho(x, y)$ piuttosto che come un parametro costante. Le due equazioni del sistema risultano pertanto non lineari, per cui l'unicità della soluzione non è più assicurata. Per recuperare l'unicità si può aggiungere una terza immagine I_3 , presa con direzione di luce l_3 tale che l_1, l_2, l_3 non giacciono sullo stesso piano, e costruendo un sistema lineare nella maniera seguente. Sia $I = [I_1, I_2, I_3]$ il vettore colonna

dei valori di intensità delle tre immagini registrati in uno stesso punto (x,y) , e sia $N = \rho(x,y)(n_1, n_2, n_3)$ il vettore colonna rappresentante la normale esterna alla superficie moltiplicata per l'albedo sempre nello stesso punto. Sia infine L la matrice 3×3 avente come righe le direzioni l_1, l_2, l_3 della sorgente di luce. Allora si può scrivere il sistema

$$I = LN$$

la cui soluzione esiste unica

$$N = L^{-1}I$$

dato che le direzioni della sorgente luminosa sono per ipotesi non coplanari, il che assicura l'invertibilità di L .

Una volta noto il vettore N , è possibile recuperare l'albedo e la normale

$$\rho = \|N\|, \quad n = 1/\rho N$$

Da quest'ultima espressione si può notare subito che la tecnica fallisce nel caso in cui l'oggetto da ricostruire abbia dei punti neri (corrispondenti a un valore nullo dell'albedo ρ). L'ipotesi di non coplanarità delle direzioni di luce impedisce inoltre l'uso di immagini prese con luce solare in diversi momenti del giorno, poichè il percorso del sole nel cielo è quasi planare. Le immagini richieste possono essere ottenute muovendo esplicitamente la sorgente luminosa, usando più sorgenti calibrate una rispetto all'altra o facendo ruotare insieme la superficie dell'oggetto e la fotocamera, per simulare l'effetto di una sorgente di luce in movimento. Non c'è inoltre restrizione sul numero di immagini da utilizzare, ammesso che sia maggiore o uguale a tre. Se si hanno a disposizione n immagini I_1, I_2, \dots, I_n prese con n diverse direzioni di luce, si può sempre risolvere il sistema sovradeterminato

$$I = LN$$

facendo uso della tecnica della pseudo-inversa di Moore-Penrose

$$N = (L^T L)^{-1} L^T I$$

E' importante notare che le direzioni della sorgente luminosa sono sempre assunte note a priori, per cui spesso questa tecnica viene denotata come *Stereo Fotometrico Calibrato*. Ulteriori dettagli sulla teoria dello Stereo Fotometrico possono essere reperiti nel lavoro in [12]

3.2 Filtraggio di Ombre Proiettate e Riflessioni speculari

Uno dei limiti principali dello stereo fotometrico deriva dal fatto che la sua teoria di base non prende in considerazione l'eventuale presenza di ombre proiettate nelle immagini, come mostrato in figura 3.3. Queste possono manifestarsi se la superficie impressa ha delle conche o delle protuberanze (nel caso di un volto umano, il profilo del naso proietta delle ombre sulle guance, mentre le arcate sopraccigliari possono mettere in ombra parte degli occhi). In queste situazioni alcuni punti della superficie risultano oscurati in almeno un'immagine, e l'inclusione di queste immagini nel sistema lineare dello stereo fotometrico può



Figura 3.3: A sinistra, immagine di un volto in cui si può osservare l'ombra proiettata dal profilo del naso. A destra, immagine di un volto contenente un gran numero di specularità.

determinare delle anomalie nelle normali e negli albedo ricavati.

Avendo a disposizione un gran numero di immagini per individuo grazie al database di Yale, i punti di un volto che possono essere soggetti a ombre proiettate appaiono effettivamente oscurati solo in alcune di esse. Una volta individuate e scartate nel calcolo del sistema, rimarrebbe comunque un numero di immagini abbastanza alto da garantire una ricostruzione accurata.

Un'altra limitazione è data dalla presenza di specularità (o *highlights*) in alcune zone della superficie, di cui viene mostrato un esempio in figura 3.3.

Si pone dunque il problema di filtrare in ogni punto il set di immagini per eliminare quelle affette da occlusioni e riflessioni speculari.

3.2.1 Filtro d'Intensità

Una tecnica di filtraggio semplice ed efficace, proposta nel lavoro in [3], sfrutta il cosiddetto *errore fotometrico* ε_p per individuare le immagini in cui un punto risulta in ombra. Per errore fotometrico si intende la somma su tutte le immagini delle differenze in valore assoluto tra intensità registrate nel punto immagine e quelle simulate dal modello lambertiano con la normale e l'albedo stimati

$$\varepsilon_p = \sum_j |I_j(x, y) - \rho n(x, y) \cdot l_j|.$$

Quando il set di immagini ne contiene un numero abbastanza grande in cui il punto è in ombra, l'errore fotometrico corrispondente risulta in genere più elevato di quello calcolato su un punto che non è mai in ombra. L'algoritmo alla base del filtro è dunque il seguente:

1. Effettuare una prima ricostruzione del punto utilizzando tutte le immagini a disposizione, ottenendo una stima iniziale dell'albedo ρ_0 e della normale alla superficie n_0 .
2. Utilizzare la normale e l'albedo calcolati per determinare l'errore fotometrico ε_p
3. Se $\varepsilon_p \leq \alpha$, il punto viene considerato non affetto da ombre proiettate, poichè le intensità simulate sono molto simili a quelle reali, e l'algoritmo

termina. Se $\varepsilon_p \geq \alpha$, ovvero se le intensità simulate sono troppo diverse da quelle reali, viene scartata l'immagine in cui il punto appare più scuro e si ritorna al punto 1. Il threshold α utilizzato nell'algoritmo è comune a tutti i punti, e va determinato in maniera sperimentale.

L'implementazione del filtro comporta un allungamento sensibile dei tempi di esecuzione dello stereo fotometrico, in quanto per alcuni punti il sistema lineare viene ricalcolato e risolto più volte prima di terminare la ricostruzione. Inoltre, per via della sua struttura, il filtro non tiene conto della presenza di specularità nelle immagini, che devono dunque essere trattate con un altro strumento.

3.2.2 Filtro Fotometrico

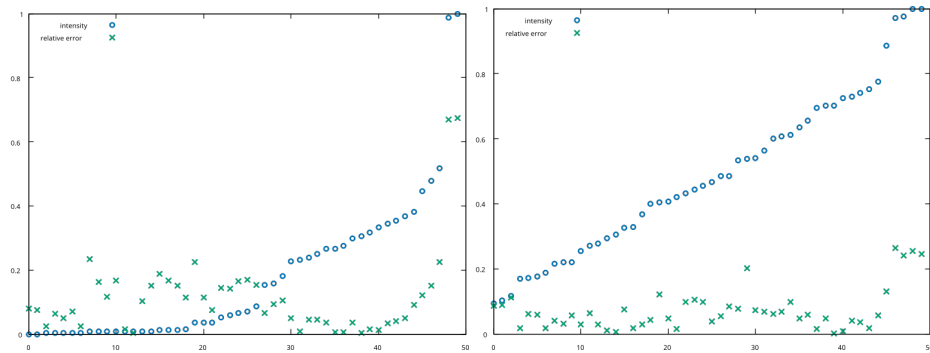


Figura 3.4: Grafici dell'errore fotometrico misurato su due vertici al variare delle intensità rilevate: gli errori maggiori corrispondono alle immagini in cui i vertici sono in ombra o presentano delle specularità.

Per riuscire a filtrare contemporaneamente specularità e ombre proiettate nel set di immagini utilizzate per ricostruire un punto, viene qui proposto un algoritmo simile al precedente, che ha come base l'individuazione delle immagini da scartare utilizzando l'errore fotometrico. L'idea è infatti che anche le specularità, come le ombre, appaiano in un numero relativamente basso di immagini (tutte quelle in cui la normale uscente cade in un piccolo intorno della bisettrice dell'angolo tra la luce e il punto di vista), per cui l'errore fotometrico "locale" ε_j , calcolato sulle singole immagini

$$I_j(x, y) - \rho n(x, y) \cdot l_j,$$

risulta elevato sia quando il punto appare in ombra che quando il punto esibisce una riflessione speculare. I grafici in figura 3.4 mostrano gli errori fotometrici locali relativi alle diverse intensità registrate nelle immagini, presi su due punti del volto (la punta e l'estremità sinistra del naso) affetti sia da ombre proiettate che da specularità.

I passi del filtro proposto sono i seguenti:

1. Effettuare una prima ricostruzione del punto utilizzando tutte le immagini a disposizione, ottenendo una stima iniziale dell'albedo ρ_0 e della normale alla superficie n_0 .

2. Utilizzare la normale e l'albedo calcolati per determinare gli errori fotometrici locali ε_j , relativi a tutte le immagini I_j .
3. Scartare dal sistema tutte le immagini I_j in cui si ha $\varepsilon_j \geq \alpha$, mantenendone sempre un numero maggiore o uguale a 3 (indispensabile per l'applicabilità dello stereo fotometrico).

Questo filtro ha il vantaggio di essere più leggero del precedente in termini di tempi di esecuzione, poichè la scrematura delle immagini viene fatta una volta per tutte dopo il primo calcolo degli errori fotometrici.

Capitolo 4

Segmentazione dei Volti



Figura 4.1: Punti chiave di un volto riconosciuti dalla libreria *dlib*.

Questo capitolo dedica un piccolo spazio alla descrizione del processo di segmentazione e discretizzazione dei volti contenuti nelle immagini. Una delle difficoltà principali che si incontrano nell'implementazione dello stereo fotometrico è infatti proprio la segmentazione del volto da ricostruire, ovvero l'individuazione del suo contorno, per evitare di computare normali ed albedo di punti esterni ad esso.

L'approccio scelto per questo lavoro è relativamente originale rispetto ai metodi più comuni usati fino ad oggi nel campo.

In genere, quando si ha a disposizione una serie di immagini del volto di un individuo che si vuole ricostruire, la parte di immagine contenente il volto viene ritagliata (molto spesso manualmente) e si definisce su di essa una griglia rettangolare regolare, avente come vertici le coordinate dei singoli pixel.

Questo modo di procedere risulta scomodo e dispendioso quando si vuole utilizzare un gran numero immagini, di dimensioni anche variabili tra loro. Il ritaglio dev'essere inoltre effettuato nella parte interna dei volti, per evitare di includere punti del resto della scena impressa.

In questa ricerca si è scelto di utilizzare una tecnica di riconoscimento facciale

introdotta di recente, che consente di automatizzare il processo di segmentazione ed ottenere delle griglie che ricoprono con precisione tutti i volti da ricostruire. Si tratta della libreria pubblica di riconoscimento facciale *dlib*, reperibile in [10], che implementa i risultati del lavoro in [8]. Questa libreria è in grado di individuare in un'immagine e restituire, sotto forma di coordinate dei pixel, un certo numero di particolari punti chiave in un volto (contorno del viso, degli occhi, del naso e delle labbra), che vengono presi come base per la costruzione di una griglia sul volto. L'algoritmo è estremamente veloce, ed è in grado di tracciare in tempo reale i contorni di volti contenuti in un video, proveniente ad esempio da una webcam.

La versione della libreria utilizzata è stata allenata a trovare 68 punti chiave, mostrati in figura 4.1.

4.1 Costruzione delle Griglie

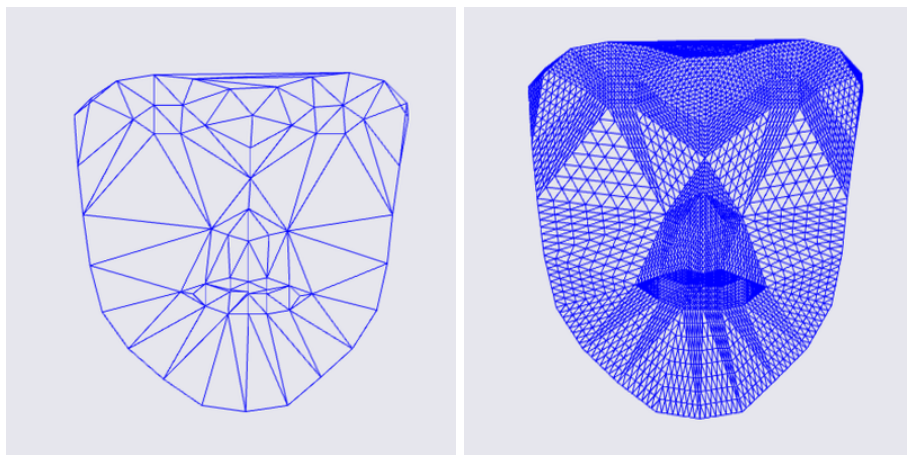


Figura 4.2: A sinistra, griglia base costruita triangolando i punti chiave di un volto. A destra, griglia raffinata risultante dall'applicazione di tre iterazioni dell'algoritmo di suddivisione.

Una volta noti i punti, ne viene effettuata una triangolazione di Delaunay che produce una griglia non regolare composta di 103 triangoli, mostrata in figura 4.2. La griglia prodotta viene successivamente raffinata, in modo da coprire il maggior numero possibile di pixel dell'immagine. Il processo di raffinamento avviene applicando a tutti i triangoli un semplice algoritmo di suddivisione che agisce come segue: considerato un triangolo, si calcolano i punti medi degli archi che lo costituiscono e, se non sono stati già stati calcolati suddividendo un triangolo vicino, si aggiungono alla lista dei vertici della griglia. La lista dei triangoli viene a questo punto aggiornata eliminando il triangolo originale ed inserendo i quattro sotto-triangoli prodotti collegando i punti medi degli archi. Ad ogni passo dell'algoritmo di suddivisione, la griglia prodotta ha dunque un numero di triangoli quattro volte superiore a quella del passo precedente.

4.2 Vantaggi Rispetto alle Griglie Tradizionali



Figura 4.3: Esempi di scambio di texture tra due griglie appartenenti a due individui diversi.

La griglia costruita triangolando i punti chiave del volto si è rivelata essere una struttura flessibile e molto potente durante la ricerca, offrendo diversi vantaggi rispetto alle griglie regolari tradizionali. Innanzitutto, la forma della griglia ricalca naturalmente quella del volto su cui viene costruita, senza bisogno di ulteriori interventi manuali. In secondo luogo, l'individuazione dei punti chiave consente di etichettare ogni vertice con la parte di volto a cui appartiene: si può tenere traccia ad esempio del gruppo di vertici che compongono l'occhio destro, della bocca interna o delle labbra. Queste informazioni possono risultare preziose in futuro per animare la griglia, e renderne controllabili le espressioni. Infine, dato che la differenza tra due griglie risiede solo nella posizione dei punti chiave, griglie di individui differenti possono scambiare tra di loro informazioni, come una normale od una coordinata immagine, e si possono creare dei volti originali mescolando ad esempio la forma di un volto con i colori di un altro (alcuni esempi in figura 4.3).

Capitolo 5

Integrazione delle Normali

Una volta calcolate le normali alla superficie di un volto con lo Stereo Fotometrico, sorge il problema di determinare a partire da queste i valori di profondità $z(x, y)$ per ottenere una ricostruzione completa.

Si consideri per semplicità di aver discretizzato la superficie da ricostruire con una griglia regolare XY , e di avere a disposizione per ogni vertice (x, y) della griglia un vettore normale $n(x, y)$, che è della forma

$$n(x, y) = n(p, q) = \frac{1}{\sqrt{1 + p^2 + q^2}}(-p, -q, 1),$$

dove $p = \frac{\partial z}{\partial x}$, $q = \frac{\partial z}{\partial y}$ sono le componenti del gradiente della funzione di profondità $z(x, y)$ della superficie. Questa relazione tra normale e gradiente può essere sfruttata per recuperare la funzione z . Nota una normale $n(x, y) = (n_x, n_y, n_z)$, le componenti del gradiente di z sono date da

$$p = -\frac{n_x}{n_z}, \quad q = -\frac{n_y}{n_z}.$$

Scrivendo dunque p e q in termini della profondità z utilizzando un rapporto incrementale

$$p = z(x + 1, y) - z(x, y), \quad q = z(x, y + 1) - z(x, y)$$

si ottiene un sistema lineare di equazioni che può essere risolto per recuperare i valori di z . Quest'approccio è risultato difficilmente praticabile. Innanzitutto per costruire il sistema lineare si ha bisogno di una griglia regolare, facilmente traducibile in una matrice, mentre quella costruita segmentando i volti con i punti chiave non lo è. In secondo luogo, per essere risolvibile il sistema ha bisogno di condizioni al bordo di tipo Neumann o Dirichlet che richiedono ulteriori informazioni sulla soluzione cercata. Le condizioni di tipo Neumann in particolare sono difficili da imporre su domini di forma più complessa di un rettangolo, come è il caso dei volti segmentati, e aggiungono ulteriori difficoltà all'implementazione. Infine, il problema principale di questa integrazione diretta è dato dal fatto che le normali ottenute dallo Stereo Fotometrico sono calcolate in maniera indipendente l'una dall'altra. Ciò può determinare delle inconsistenze piuttosto gravi quando si prova ad integrarle, poichè le immagini utilizzate per recuperarle potrebbero contenere del rumore, andando così ad intaccare il vincolo di

integrabilità

$$\frac{\partial p}{\partial y}(x, y) = \frac{\partial q}{\partial x}(x, y)$$

Per superare questi inconvenienti sono state sviluppate diverse tecniche basate sul calcolo delle variazioni, come il metodo di Horn ([7]), che cerca la funzione z che minimizza il funzionale

$$\int \int (z_x - p)^2 + (z_y - q)^2 dx dy$$

la cui equazione di Eulero-Lagrange associata è data da

$$\Delta z = p_x + q_y.$$

Di nuovo, anche questa equazione ha bisogno di condizioni al contorno per essere risolvibile, condizioni date nello specifico da

$$(z_x, z_y) \cdot n_c = (p, q) \cdot n_c$$

dove n_c indica la normale esterna al dominio, ed è complicata da discretizzare e risolvere su griglie non strutturate e domini di forma complessa.

5.1 Equazione Eikonale e Schemi Semi-Lagrangiani

Una maniera alternativa per recuperare i valori di profondità dalle normali consiste nel porre l'attenzione sul modulo gradiente della superficie, piuttosto che sulle sue singole componenti. Note le normali nei vertici, e dunque le componenti p e q del gradiente in ogni punto, la funzione di profondità $z(x, y)$ cercata soddisfa l'equazione

$$|\nabla z| = \sqrt{p^2 + q^2}.$$

Quest'equazione è nota col nome di *equazione eikonale*, e compare in numerosi ambiti di ricerca, dalla computer grafica alla dinamica granulare.

Il vantaggio principale che deriva da questo approccio al problema di integrazione risiede nel fatto che per risolvere l'equazione eikonale sono stati sviluppati dei metodi numerici particolarmente adatti ad essere implementati su griglie non strutturate. Si tratta degli schemi Semi-Lagrangiani, che nascono dall'idea di discretizzare il modulo del gradiente della funzione cercata nella forma di derivata direzionale. Più precisamente, il modulo del gradiente in un punto è dato dalla derivata direzionale massima della funzione nel punto, ovvero

$$|\nabla z| = \max_{a \in B(0,1)} \left(\frac{z(x) - z(x - ah)}{h} \right).$$

Sostituendo questa espressione nell'equazione eikonale e riordinando i termini si ottiene il seguente schema di punto fisso

$$z(x) = \min_{a \in B(0,1)} (z(x - ah)) + h\sqrt{p^2 + q^2}$$

Per tradurlo su una griglia discreta si può partire da uno stato iniziale z_0 , scegliere per ogni vertice della griglia un certo numero di direzioni e un passo h

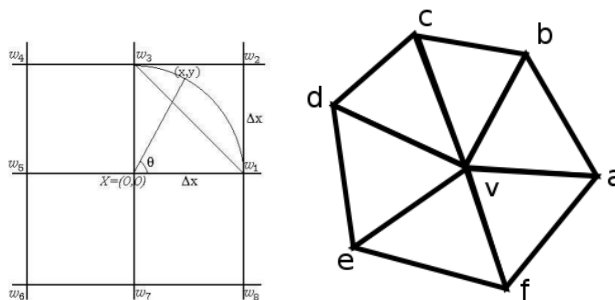


Figura 5.1: A sinistra, direzioni per lo schema semi-lagrangiano prese su una griglia regolare. A destra, con la griglia non regolare le direzioni scelte sono quelle che puntano ai vicini.

ed andare alla ricerca del minimo della funzione lungo le direzioni fissate, applicando lo schema iterativamente su tutti i vertici fino a convergenza.

Nel caso di una griglia regolare, per semplicità le direzioni e il passo vengono scelti in maniera da cadere sui vicini $z_{i+1,j}$, $z_{i-1,j}$, $z_{i,j+1}$, $z_{i,j-1}$ del vertice $z_{i,j}$. Ulteriori direzioni, in genere quelle intermedie, possono essere aggiunte a piacere per aumentare la precisione dello schema, con lo svantaggio di dover interpolare i valori della funzione nelle nuove direzioni.

L'applicazione di questo schema ad una griglia non strutturata risulta piacevolmente semplice. In una griglia del genere, ogni vertice è connesso ad un certo numero, non necessariamente sempre lo stesso, di vertici vicini. In particolare, nella griglia costruita con i punti chiave di un volto e suddivisa come mostrato nel capitolo precedente, ogni vertice interno possiede almeno sei vertici vicini, corrispondenti a sei direzioni diverse. Si tratta di un numero abbastanza elevato da permettere l'applicazione dello schema semi-lagrangiano senza eccessiva perdita di precisione. Detto \mathfrak{N} l'insieme dei vertici vicini ad un certo vertice z_j , lo schema semi-lagrangiano per questo vertice è dato da

$$z_j = \min_{z \in \mathfrak{N}} z + h\sqrt{p^2 + q^2}$$

dove il passo h viene scelto ad ogni iterazione come la distanza sulla griglia tra il vertice z_j e il vertice minimo tra i suoi vicini in \mathfrak{N} .

Per quanto riguarda le condizioni da applicare sul bordo, la geometria dei volti umani dà l'idea che una scelta sensata sia data dall'imposizione dei cosiddetti *vincoli di stato*. Avvicinandosi ai punti sul bordo del viso, il profilo della testa risulta quasi perpendicolare al piano dell'immagine. Possiamo pensare che in questi punti il modulo del gradiente alla superficie sia infinito (da cui il vincolo), pertanto su di essi la ricerca del vicino minimo non avrà alcun effetto nell'applicazione dello schema di punto fisso. In conclusione sui punti di bordo basta cercare il minimo nelle direzioni che puntano all'interno del volto, direzioni sempre disponibili, in quanto il dominio definito dalla griglia è un insieme convesso e fortemente connesso. Dal punto di vista implementativo dunque, i punti di bordo non hanno bisogno di attenzioni particolari rispetto ai punti interni.

A dispetto della semplicità di discretizzazione, l'equazione eikonale è carat-

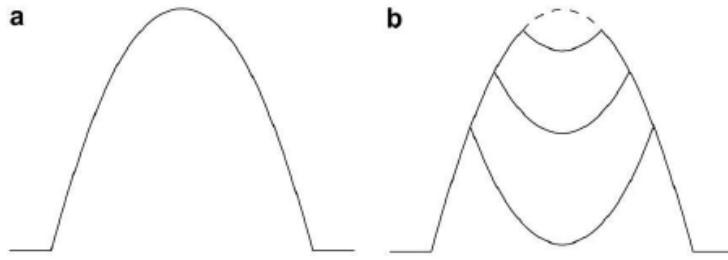


Figura 5.2: Effetti dell'ambiguità concavo-convesso dell'equazione eikonale: le nuove funzioni generate dalle riflessioni sono tutte soluzioni della stessa equazione.

terizzata da un'ambiguità di risoluzione, dovuta al fatto che le componenti del gradiente delle sue soluzioni sono uguali tra di loro a meno di un segno. Dato che l'equazione viene imposta sul modulo del gradiente, questa non effettua infatti distinzioni tra funzioni concave e funzioni convesse. Come mostrato in figura 5.2, se la soluzione cercata ha un punto di massimo o di minimo, effettuando dei tagli ad una quota qualsiasi e invertendo la convessità della funzione se ne ottiene un'altra che è ancora soluzione dell'equazione eikonale.

Si tratta di un inconveniente piuttosto scomodo, che costringe a fissare nello stato iniziale z_0 i valori di tutti i punti critici della funzione.

5.2 Implementazione Fast-Marching

Il vincolo di dover fissare i valori dei punti critici della soluzione dell'equazione eikonale suggerisce un'idea interessante per accelerare la convergenza dello schema di punto fisso semi-lagrangiano. Una volta fissato il valore di un minimo, i valori dei suoi vicini vengono determinati dallo schema di punto fisso in una sola iterazione. Questi vertici andranno infatti a cercare ad ogni iterazione il loro vicino minimo, e questo sarà sempre lo stesso valore fissato inizialmente.

Se la funzione cercata possiede un unico minimo globale, partendo da questo e procedendo tra i suoi vicini, si può fissare sequenzialmente il valore di ogni altro vertice, dal più piccolo al più grande, senza dover iterare lo schema su tutta la griglia più volte.

Questa idea è la base della tecnica più generale *Fast Marching*, introdotta da Sethian in [13] come tecnica per ottimizzare i metodi level-set.

La strategia proposta da Sethian può essere riassunta nei seguenti punti

1. Fissare il valore dei punti di minimo della soluzione, che vanno a costituire il fronte iniziale e vengono etichettati come punti "Vivi". Inizializzare il valore di tutti i restanti punti ad infinito (o ad un numero ragionevolmente grande) ed etichettarli come "Lontani".
2. Effettuare un'iterazione dello schema semi-lagrangiano su tutti i vertici vicini a quelli del fronte iniziale. Questi vertici vengono etichettati come "Vicini" e costituiscono la cosiddetta *Narrow Band*.
3. Scegliere tra i punti "Vicini" il più piccolo, fissarne il valore ed etichettarlo come punto "Vivo".

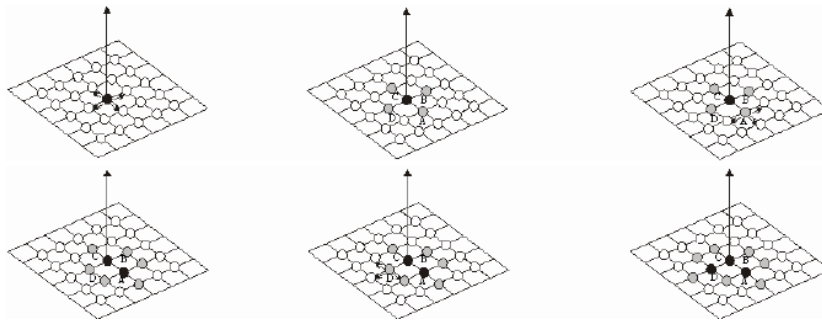


Figura 5.3: Schema grafico del primo passo dell'algorithm Fast Marching: il vertice Vivo appartenente al fronte iniziale aggiorna il valore di tutti i suoi vicini, che vanno a costituire la Narrow Band. Tra i vertici della Narrow Band viene scelto il più piccolo e ne viene fissato il valore, aggiornando al contempo i valori dei suoi vertici vicini.

4. Applicare un'iterazione dello schema semi-lagrangiano a tutti i vicini non "Vivi" del punto appena fissato ed etichettare come "Vicini" tutti i vicini che sono ancora "Lontani"
5. Tornare al punto 3 finchè tutti i vertici non risultano "Vivi".

Si può osservare in figura 5.3 un passo dell'algorithm, ed in figura 5.4 una rappresentazione grafica della Narrow Band.

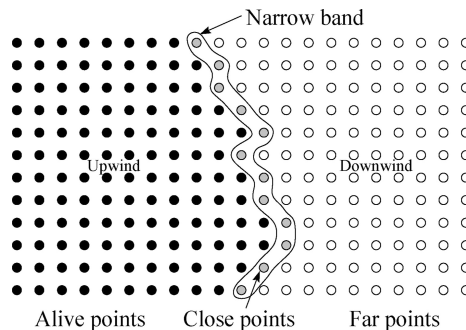


Figura 5.4: Rappresentazione della Narrow Band in uno stato intermedio dell'algorithm Fast Marching.

L'implementazione della tecnica di Sethian risulta particolarmente vantaggiosa dal punto di vista computazionale rispetto allo schema originale iterativo. Per una griglia composta di N vertici, una scelta intelligente della struttura dati che tiene traccia dei punti della Narrow Band (come ad esempio un albero rosso e nero) consente di trovarne il minimo in tempo ($O(\log N)$). Dato che ogni punto viene visitato una sola volta, il costo computazionale totale dell'algorithm è di ordine ($O(N \log N)$). In applicazioni reali ([2]) si è osservato che il tempo impiegato dall'algorithm Fast Marching per costruire la soluzione a parità di precisione è circa due ordini di grandezza più piccolo di quello impiegato dallo schema iterativo per arrivare a convergenza.

5.3 Strategia di Penalizzazione

Per applicare il Fast Marching alla ricostruzione dei volti resta dunque il problema di determinare il valore della superficie cercata sui suoi punti critici. Questi non sono in generale noti a priori, ma supponendo che ce ne sia solo uno, il suo valore può essere imposto a piacimento, dato che la superficie ricostruita dalle normali è unica a meno di una traslazione spaziale. Un'osservazione interessante che si può fare a questo punto è la seguente: considerando un volto con lo sguardo rivolto verso la direzione negativa dell'asse Z , il suo punto di minimo globale cadrà sulla punta del naso (come mostrato in figura). I punti chiave individuati dalla libreria *dlib* per costruire la griglia contengono anche la posizione della punta del naso, fornendo un'informazione molto preziosa sul punto da cui far partire il Fast Marching.

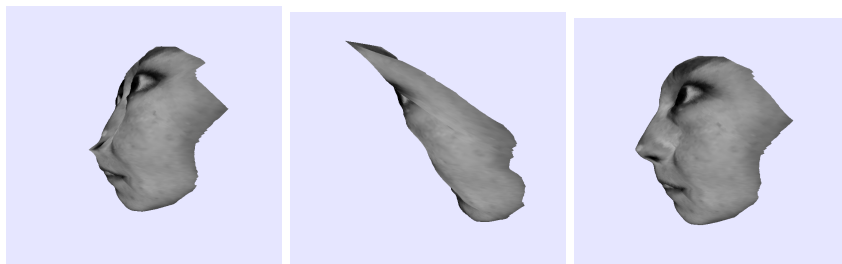


Figura 5.5: Superfici ottenute facendo partire l'algoritmo fast marching da vertici diversi. A sinistra, da un vertice alla base di una narice. Al centro, da un vertice del sopracciglio sinistro. A destra, da un vertice sulla punta del naso.

La figura 5.5 mostra alcuni risultati ottenuti inizializzando il fronte iniziale in punti differenti, e risulta evidente come la scelta della punta del naso risulti quella che dà risultati migliori. Anche in questo caso però i risultati non sono soddisfacenti: la forma ottenuta risulta eccessivamente convessa, proprio per via del fatto che alcuni minimi locali della superficie impediscono il corretto funzionamento del Fast Marching.

Per risolvere l'inconveniente si è fatto riferimento all'idea di Kriegman, riportata in [6], di risolvere l'equazione eikonale per una funzione w della forma

$$w = z + \lambda f$$

invece che direttamente per z , dove λ è una costante positiva ed f è una funzione scelta in modo da avere un unico punto critico, coincidente con il minimo globale di z . Il ruolo del termine λf è quello di cancellare tutti i punti critici della funzione w all'infuori del minimo globale condiviso con z .

Una scelta possibile per la funzione f , suggerita da Kriegman, è quella di porla uguale alla funzione distanza al quadrato dal punto di minimo globale. Supponendo che questo minimo cada in un certo punto della griglia (x_0, y_0) , si può scrivere

$$f(x, y) = (x - x_0)^2 + (y - y_0)^2,$$

e l'equazione eikonale corrispondente da risolvere diventa

$$|\nabla w| = \sqrt{(p + 2\lambda(x - x_0))^2 + (q + 2\lambda(y - y_0))^2}.$$

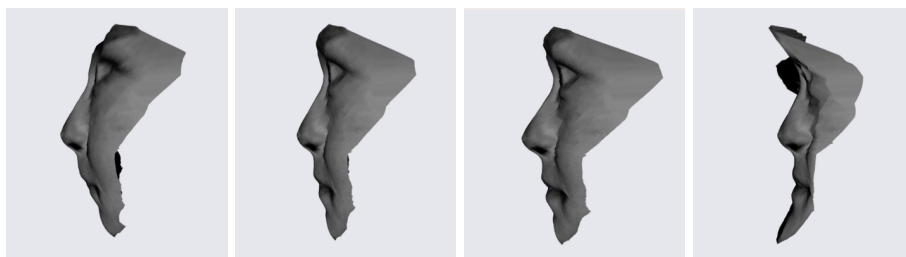


Figura 5.6: Effetti del parametro di penalizzazione λ sull'integrazione delle normali: da sinistra a destra $\lambda = 0, 0.5, 0.7, 2$.

Come mostrato da Kriegman, scegliere un λ abbastanza grande assicura che la funzione w risulti priva di altri punti critici all'infuori del minimo globale, e una volta nota w con il Fast Marching si può recuperare la funzione z cercata in origine ponendo

$$z = w - \lambda f.$$

Occorre tuttavia fare attenzione nel fissare il valore di λ , in quanto se lo si sceglie troppo grande, allontanandosi dal punto iniziale il termine $\lambda \nabla f$ diventa preponderante nell'equazione eikonale, corrompendo il risultato finale a causa di errori di arrotondamento del calcolatore. Al contrario, scegliere un λ troppo piccolo può far sopravvivere alcuni punti critici di w , minando così l'applicabilità del Fast Marching. L'immagine 5.6 mostra gli effetti della scelta di λ sulla ricostruzione di un volto.

5.3.1 Stima del Parametro di Penalizzazione



Figura 5.7: Immagini renderizzate con Blender di due dei volti artificiali utilizzati per stimare il parametro di penalizzazione.

La questione di determinare una scelta ottimale del parametro di penalizzazione λ è stata affrontata facendo uso di alcuni volti artificiali, costruiti con l'aiuto del programma MakeHuman (reperibile gratuitamente da [14]) e di cui si possono osservare degli esempi in figura.

I volti sono stati esportati nel software di animazione Blender (scaricabile gratuitamente da [5]) ed illuminati con delle sorgenti luminose direzionali poste in diverse direzioni, producendo delle immagini artificiali che possono essere usate per applicare lo Stereo Fotometrico e ottenere le normali dei volti.

Le superfici ricostruite integrando le normali con la tecnica Fast Marching sono

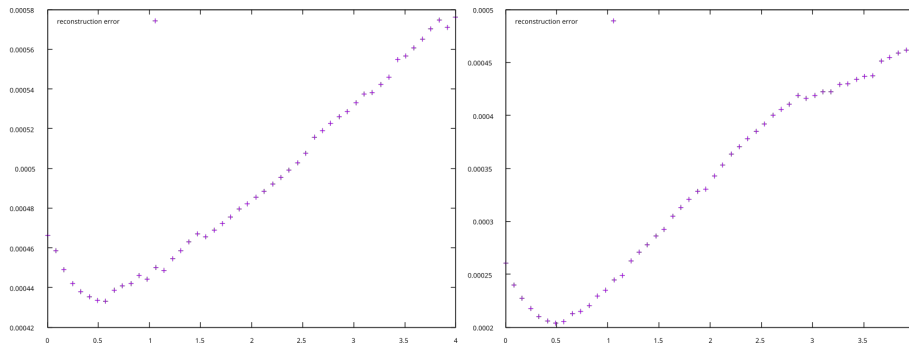


Figura 5.8: Grafico dell'errore di ricostruzione medio al variare del parametro λ per due volti artificiali.

state infine confrontate con i volti originali, in modo da determinare l'errore commesso sui valori di profondità calcolati a partire dalle immagini.

Osservando la risposta dell'errore medio alla variazione del parametro di penalizzazione λ applicato, è emerso un trend interessante, che può essere osservato in figura 5.8. Si è visto infatti che il valore di λ corrispondente al minimo errore è all'incirca 0.5 per tutti i volti ricostruiti.

Ulteriori dettagli sullo script sviluppato per produrre i rendering artificiali sono riportati in [4].

Capitolo 6

Risultati

6.1 Banco di Prova: Il Database di Yale

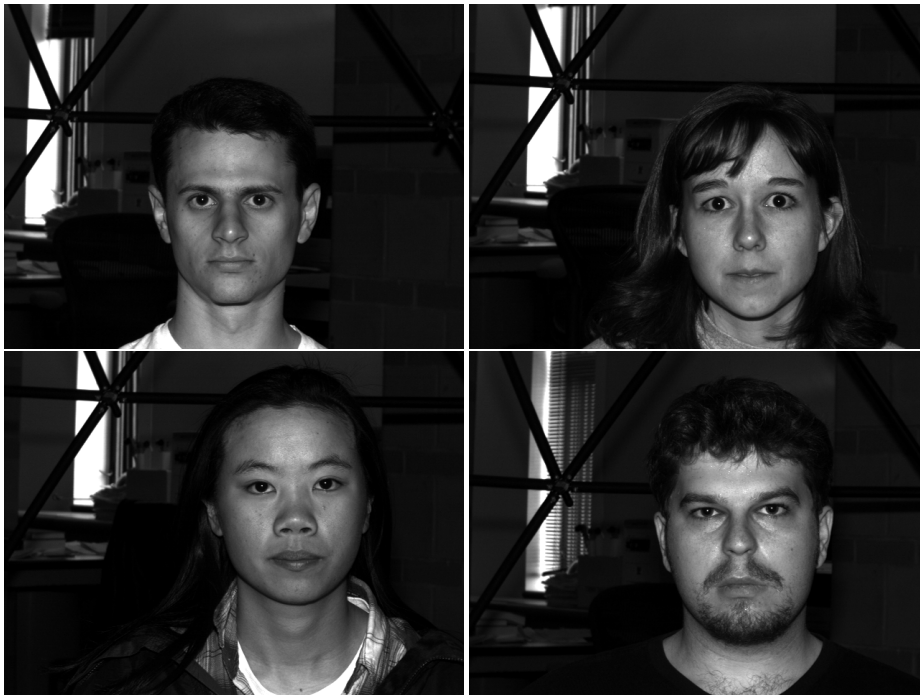


Figura 6.1: Alcuni volti del database di Yale.

Il database di volti dell'università di Yale contiene un totale di 16128 immagini di 28 soggetti umani, prese con 9 pose differenti e 64 diverse direzioni della sorgente luminosa. Si tratta di uno dei database più completi disponibili al pubblico, ed è quello maggiormente utilizzato per testare nuove tecniche di ricostruzione e di riconoscimento facciale.

Le immagini, fornite in toni di grigio con il formato .pgm, hanno tutte la stessa risoluzione 640×480 , e sono state catturate con una fotocamera Sony XC-75



Figura 6.2: Volto del database di Yale fotografato con diverse direzioni della luce.

utilizzando una cupola di illuminazione costruita per l'occasione. La cupola è attrezzata con 64 luci stroboscopiche regolate da un computer, ed è in grado di catturare le 64 immagini in poco meno di 2 secondi (con una media di scatti di circa 30 frames al secondo).

La posa del volto e la direzione della sorgente luminosa in ogni immagine sono contenute nel nome del file corrispondente. Come esempio, l'immagine

`'yaleB11.P00A+35E+40.pgm'`

appartiene al soggetto numero 11 visto in posa frontale (P00), e la direzione della luce rispetto agli assi della fotocamera è a 35 gradi in azimuth (A+35) e 40 gradi in elevazione (E+40). Un azimuth positivo implica che la sorgente emette luce dalla destra dell'individuo, mentre un azimuth negativo denota una sorgente posizionata a sinistra. Valori positivi o negativi dell'angolo di elevazione denotano rispettivamente una sorgente posta al di sopra o al di sotto dell'orizzonte.

Il database viene accompagnato da un file contenente le direzioni della sorgente luminosa in coordinate sferiche. Questo file al momento non è più disponibile per via di lavori di manutenzione sul server dell'università di Yale, per cui è stato scritto un semplice script in Python (reperibile in [4]) che consente di recuperare le coordinate leggendo i nomi dei file del database.

Le immagini selezionate inizialmente per la ricostruzione di ogni individuo sono quelle con posa del volto frontale e con angoli di azimuth ed elevazione minori di 90 gradi in valore assoluto (per evitare di includere immagini con sorgenti

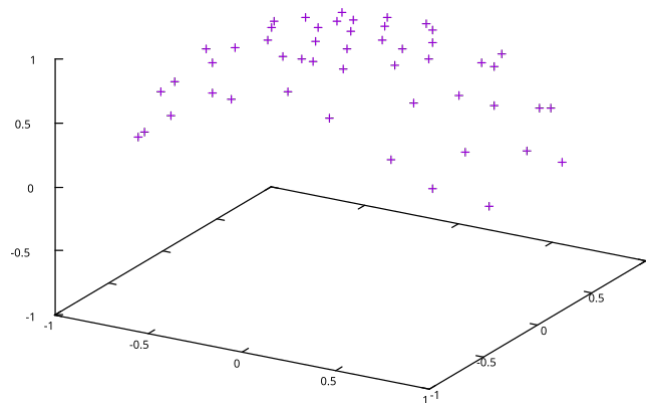


Figura 6.3: Posizioni della sorgente luminosa utilizzata per produrre il database di Yale.

luminose poste dietro il volto) per un totale di 50 immagini per individuo.

Un'ulteriore scrematura è stata svolta seguendo i risultati del lavoro in [11], che si è occupato di studiare la riflettanza della pelle umana. E' stato mostrato infatti che sotto certe angolazioni di luce incidente, la pelle esibisce proprietà di riflettanza quasi lambertiana. Nello specifico, l'analisi ha mostrato che per angoli di incidenza della luce intorno ai 30 gradi la riflettanza è ben approssimabile dal modello lambertiano, con deviazioni sensibili osservate per angolazioni superiori ai 60 gradi.

Sulla base di queste osservazioni si è scelto di limitare l'angolo di incidenza al valore medio di 45 gradi, che risulta in un totale di 26 immagini per ogni individuo.

6.2 Alcune Ricostruzioni dei Volti di Yale

Le figure mostrate in questa sezione presentano alcuni esempi di ricostruzione dei volti del database. Nella prima serie di immagini, in figura 6.4, vengono riportati i colori e le normali dei volti, rappresentate graficamente colorando ogni vertice con le componenti della sua normale come terna RGB.

La seconda serie, in figura 6.5 presenta invece le superfici prodotte dal metodo di integrazione delle normali, ottenute scegliendo il parametro di penalizzazione λ come mostrato nel capitolo precedente.

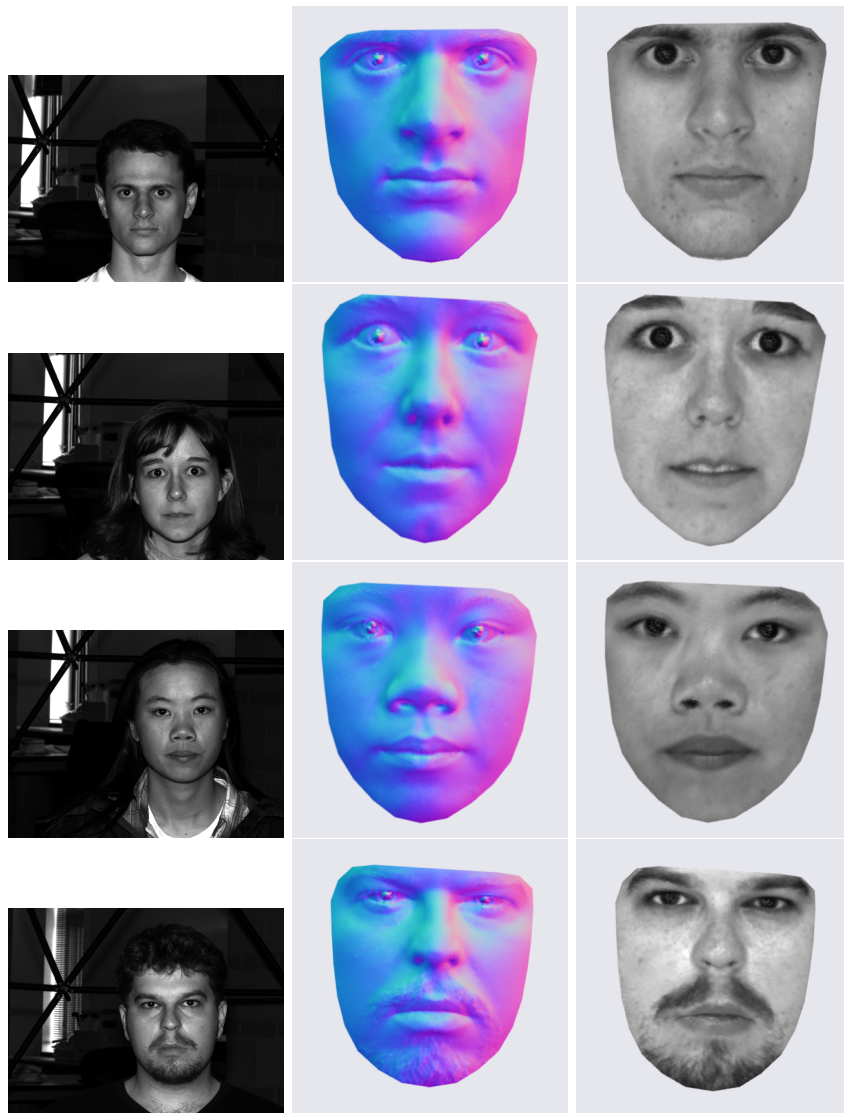


Figura 6.4: Alcuni risultati di ricostruzione di normali ed albedo dei volti con lo Stereo Fotometrico. A sinistra, immagini dei volti di riferimento. Al centro, normali alla superficie convertite in colore RGB. A destra, albedo dei volti.

6.3 Effetti del Filtraggio delle Immagini

I filtri proposti per controllare la scelta delle immagini sono stati applicati alla ricostruzione dei volti del database di Yale, e le figure mostrano graficamente il numero di immagini scartate per ogni punto al variare della scelta del threshold (colorando i vertici con una sfumatura di grigio corrispondente al rapporto tra numero di immagini utilizzate e numero di immagini totali) l'errore fotometrico commesso e l'albedo computato.

Nel processo di integrazione delle normali si è osservato che il filtro d'intensità

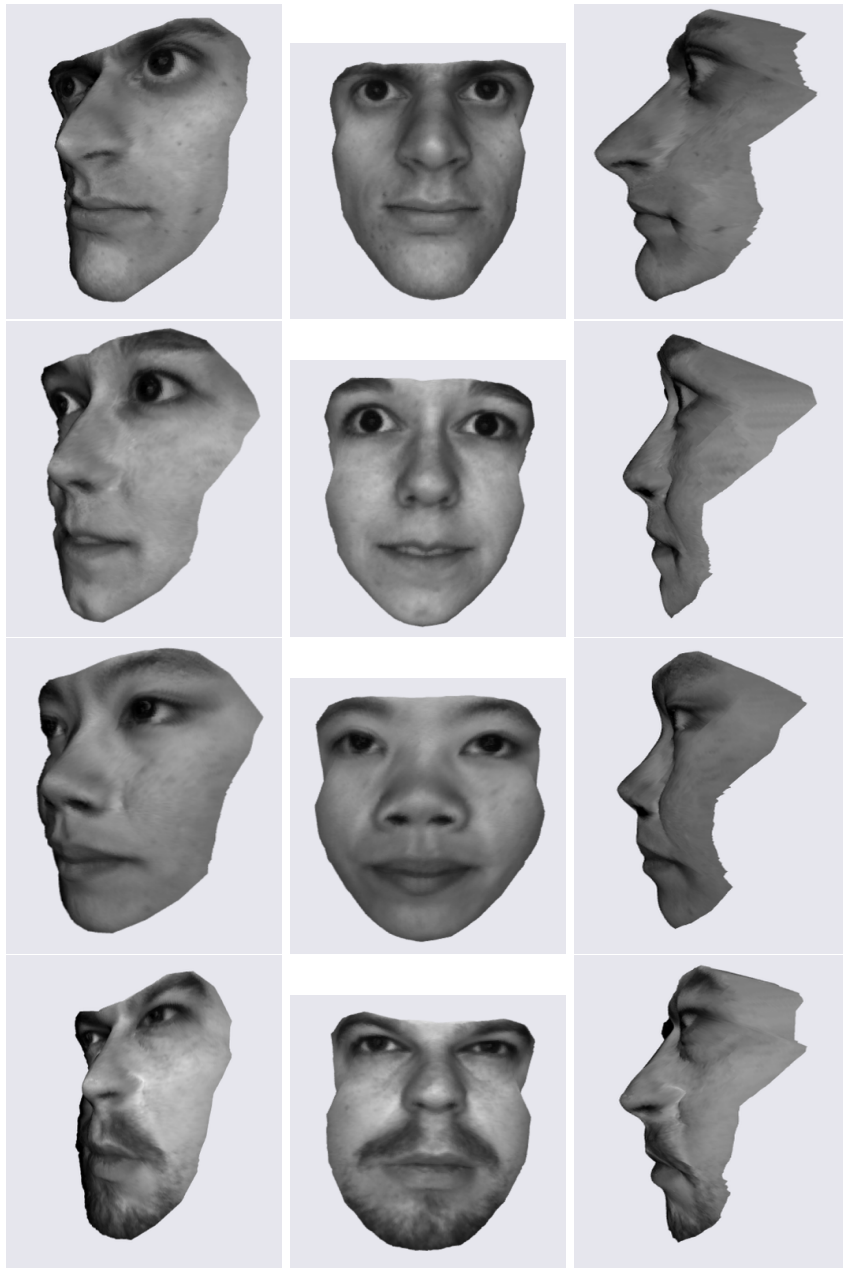


Figura 6.5: Superfici dei volti ottenute integrando le normali.

ha problemi a ricostruire punti molto scuri come le narici dei volti e l'interno degli occhi. Per le narici ciò può essere dovuto al fatto che appaiono in ombra in un numero cospicuo di immagini, e quelle sopravvissute al processo di scarto sono troppo poche per determinare una ricostruzione accurata, mentre per gli occhi i problemi principali sono sorti nella ricostruzione del contorno dell'iride, che muovendosi leggermente tra uno scatto e l'altro può dare l'impressione di

essere composta da punti in ombra piuttosto che da punti naturalmente scuri. Nel complesso entrambi i filtri sembrano dare risultati simili.



Figura 6.6: Effetti del threshold applicando il filtro di intensità alla ricostruzione dell'albedo di uno dei volti del database: da sinistra a destra threshold = 1, 0.1, 0.01. In alto: rappresentazione del numero di immagini scartate dal filtro su ogni vertice (un tono di grigio più scuro equivale a più immagini scartate per ricostruire il vertice). In basso: albedo ricostruito.

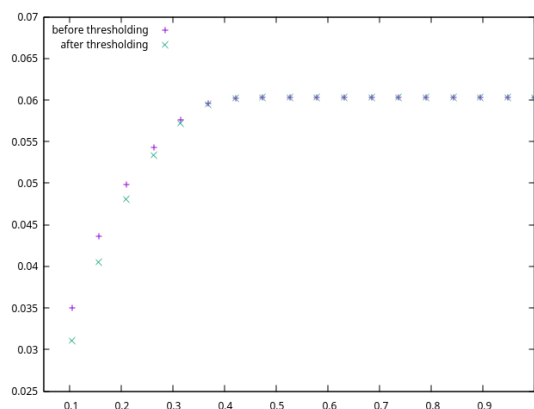


Figura 6.7: Errore fotometrico medio di una ricostruzione prima e dopo l'applicazione del filtro fotometrico, al variare del valore di thresholding.

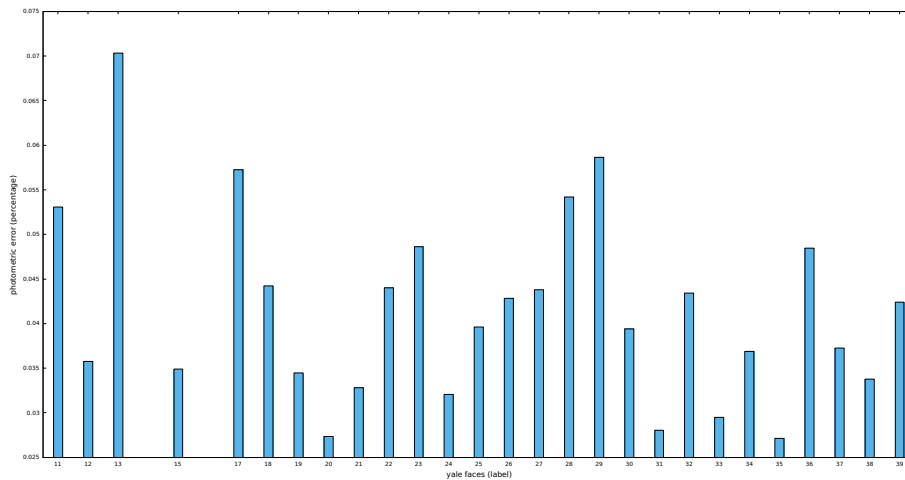


Figura 6.8: Errori fotometrici medi su tutti gli individui del database di Yale.

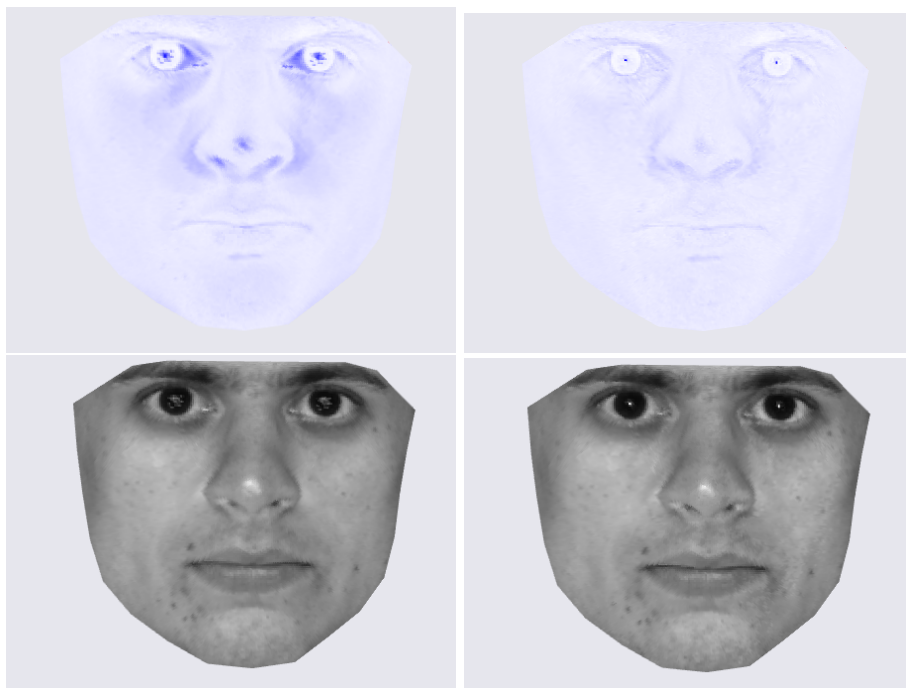


Figura 6.9: Errore fotometrico ed albedo prima (a sinistra) e dopo (a destra) l'applicazione del filtro fotometrico, con $\text{threshold} = 0.1$. Si può osservare una netta riduzione degli effetti delle specularità negli occhi dell'individuo, dovuta all'applicazione del filtro fotometrico.

6.4 Ricostruzioni da Singole Immagini

Vengono mostrati in figura 6.10 dei volti ricostruiti da singole immagini, ottenuti costruendo delle griglie dai punti chiave nelle immagini ed integrando le normali

di uno dei volti del database di Yale sui vertici delle griglie. Sulle superfici vengono poi applicate le texture dei volti originali.



Figura 6.10: Alcune ricostruzioni da singole immagini prese dal web, ottenute combinando le normali di uno dei volti del database con le griglie e le texture delle immagini.

Capitolo 7

Conclusioni e Prospettive Future

Nonostante si siano osservate alcune lievi inconsistenze nelle superfici ricostruite dai volti del database di Yale, potenzialmente dovute allo sbilanciamento tra le dimensioni dei triangoli delle griglie prodotte con i punti chiave, i risultati ottenuti sono molto soddisfacenti.

In alcuni casi si è osservato che il metodo proposto è in grado di ricostruire i dettagli più minuti di un volto, come nei e piccole cicatrici, e la progressiva diffusione di fotocamere ad alta risoluzione a prezzo di mercato sempre più accessibile consente di ottenere ricostruzioni potenzialmente molto più precise.

Il lavoro svolto getta inoltre delle basi solide per la sperimentazione di altre tecniche di ricostruzione, come lo stereo multiview. L'individuazione dei punti chiave da parte della libreria di riconoscimento permette infatti il calcolo in maniera piuttosto accurata degli angoli di posa del volto, informazione che potrebbe essere combinata allo stereo fotometrico o ad una tecnica di Shape from Shading per consentire l'utilizzo di immagini di volti non necessariamente in posa frontale.

Potrebbe essere interessante ripercorrere la strada del lavoro in [9], in cui le superfici ricostruite vengono impiegate per stimare la direzione della luce in immagini prese in ambiente non controllato. Sempre in [9], viene introdotto un nuovo modello di illuminazione, basato sulle armoniche sferiche, in grado di descrivere configurazioni di luce più generali di quella direzionale utilizzata per questa ricerca, consentendo l'applicazione dello stereo fotometrico a collezioni di immagini qualunque, con condizioni di illuminazione non necessariamente controllate, prese ad esempio tramite una webcam o da un album fotografico.

Bibliografia

- [1] Antonin Chambolle. A uniqueness result in the theory of stereo vision : coupling shape from shading and binocular information allows unambiguous depth reconstruction. *Annales de l'I.H.P. Analyse non linéaire*, 11 (1):1–16, 1994. URL <http://eudml.org/doc/78323>.
- [2] Emiliano Cristiani. *Fast Marching and Semi-Lagrangian Methods for Hamilton-Jacobi Equations with Applications*. PhD thesis, 2006.
- [3] Amina Dulac, Sandy Martedi, Hideo Saito, Kouichi Tezuka, and Masayoshi Shimizu. Combination method: photometric stereo with shadows. *Proc. 2014 Irish Machine Vision and Image Processing (IMVIP2014)*, pages 137–142, 2014.
- [4] Matteo Fois. <https://github.com/giomatfois62/faceReconstruction.git>.
- [5] Blender Foundation. Blender. <https://www.blender.org/>.
- [6] Jeffrey Ho, Jongwoo Lim, Ming-Hsuan Yang, and David Kriegman. Integrating surface normal vectors using fast marching method. pages 239–250, 2006.
- [7] Berthold KP Horn and Michael J Brooks. *The variational approach to shape from shading*, volume 33. Elsevier, 1986.
- [8] Vahid Kazemi and Josephine Sullivan. One millisecond face alignment with an ensemble of regression trees. pages 1867–1874, 2014.
- [9] Ira Kemelmacher-Shlizerman and Ronen Basri. 3d face reconstruction from a single image using a single reference face shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(2):394–405, 2011.
- [10] Davis King. dlib. <http://dlib.net/>.
- [11] Stephen R Marschner, Stephen H Westin, Eric PF Lafortune, Kenneth E Torrance, and Donald P Greenberg. Image-based brdf measurement including human skin. pages 131–144, 1999.
- [12] Roberto Mecca and Maurizio Falcone. Uniqueness and approximation of a photometric shape-from-shading model. *SIAM Journal on Imaging Sciences*, 6(1):616–659, 2013.

- [13] JA Sethian. *Fast marching methods and level set methods for propagating interfaces*, volume 3. Van Kareman Institute, 1998.
- [14] MakeHuman Team. Makehuman. <http://www.makehuman.org/>.
- [15] Robert J Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19, 1980.