

Principi ed applicazioni dell'elaborazione digitale delle immagini

di Giuseppe Cardinale Ciccotti

Un famoso aforisma recita: «Un'immagine è più significativa di diecimila parole»; questa regola, applicata fino all'inverosimile nella nostra epoca dominata dal proliferare incontrollato di messaggi ed informazioni attraverso qualsiasi canale i media possano sfruttare od inventare, ha creato in qualche misura una cultura dell'immagine che appartiene al bagaglio di ognuno di noi. «Potevamo stupirvi con effetti speciali» ormai non ha più molto significato: l'effetto speciale, quantunque sempre più sofisticato, non stupisce più nessuno, il nostro occhio si è dunque abituato alle performance dei sistemi di comunicazione

È quasi inutile affermare che il salto di qualità nella manipolazione delle immagini è avvenuta con l'avvento di macchine digitali che in breve tempo hanno scatenato la frenesia dei creativi offrendo loro una pressoché illimitata scelta di effetti e personalizzazioni.

Come si può facilmente intuire la diffusione di tecnologia a basso costo ha poi definitivamente contribuito a «fomentare» questa rivoluzione anche a livello personale ed oggi chiunque con un minimo di supporto elettronico può soddisfare la propria vena creativa.

Quello che molti lettori non sanno è che tali tecniche e tecnologie sono ormai sperimentate sin dagli anni Sessanta praticamente in concomitanza con

l'incedere della tecnologia televisiva, e che in alcuni settori come per esempio il telerilevamento, certe funzioni che oggi troviamo in qualsiasi sistema grafico, erano già ampiamente sperimentate.

L'elaborazione delle immagini è quindi un'importante branca della scienza dell'informazione e pur poggiando su di una sistematizzazione teorica ormai completa e consolidata, rimane molto dinamica nelle applicazioni e, cosa non trascurabile, assai comprensibile anche a chi vi si avvicini da profano, poiché, è proprio il caso di dirlo, i risultati sono sotto gli occhi di tutti!

In queste pagine abbiamo intenzione di mettere a fuoco le tecniche dell'elaborazione digitale delle immagini e la



Figura 1
Convenzioni adottate per l'immagine. L'origine delle coordinate è nell'angolo in alto a sinistra.

teoria a cui si rifanno. Tratteremo di trasformazioni, di filtraggi, di campionamenti, di tecniche per la codifica e la decodifica delle immagini e di tutti gli argomenti che i lettori interessati vorranno segnalargli.

Per visualizzare in pratica quello di cui tratteremo, ci avvarremo ove possibile di un pacchetto molto diffuso, Photoshop per Macintosh; tuttavia i lettori che utilizzano un qualsiasi altro tipo di sistema avranno sufficienti informazioni per riprodurre i medesimi risultati con un po' di programmazione.

La rappresentazione digitale dell'immagine

Un'immagine, che per semplicità ci limitiamo a considerare monocromatica, è definita come una funzione $f(x,y)$ bidimensionale dell'intensità luminosa, dove x e y sono le coordinate cartesiane e il valore di f assunto in ciascun punto (x,y) è proporzionale alla luminosità (al livello del grigio) in quel punto.

Un esempio che illustra le convenzioni che verranno di seguito adoperate è dato in figura 1.

Un'immagine digitale è un'immagine che è stata discretizzata in entrambe le coordinate spaziali come pure nella luminosità. Ciò significa che la funzione $f(x,y)$ continua è stata sostituita da una serie ordinata di valori rappresentativi della funzione. In definitiva l'immagine digitale non è altro che una matrice bidimensionale nella quale gli indici di riga e di colonna individuano un punto dell'immagine e l'elemento corrispondente della matrice identifica il livello di grigio di quel punto. Gli elementi di tale matrice vengono detti picture element o pixel.

La grandezza di questo array non è definita a priori e può variare con l'applicazione, tuttavia per ragioni operative, vengono scelte matrici quadrate in cui le dimensioni ed il numero dei livelli di grigio siano potenze di 2. Per esempio una risoluzione tipica per la rappresentazione di un'immagine televisiva è 512×512 pixel a 256 livelli di grigio.

Un sistema base per l'elaborazione di immagini

Descriviamo brevemente come può essere organizzato un piccolo sistema general-purpose per fare degli esperimenti di elaborazione delle immagini.

In figura 2 notate come sia presente un non meglio identificato dispositivo di acquisizione; questa unità può essere di varia natura dipendendo dall'applicazio-

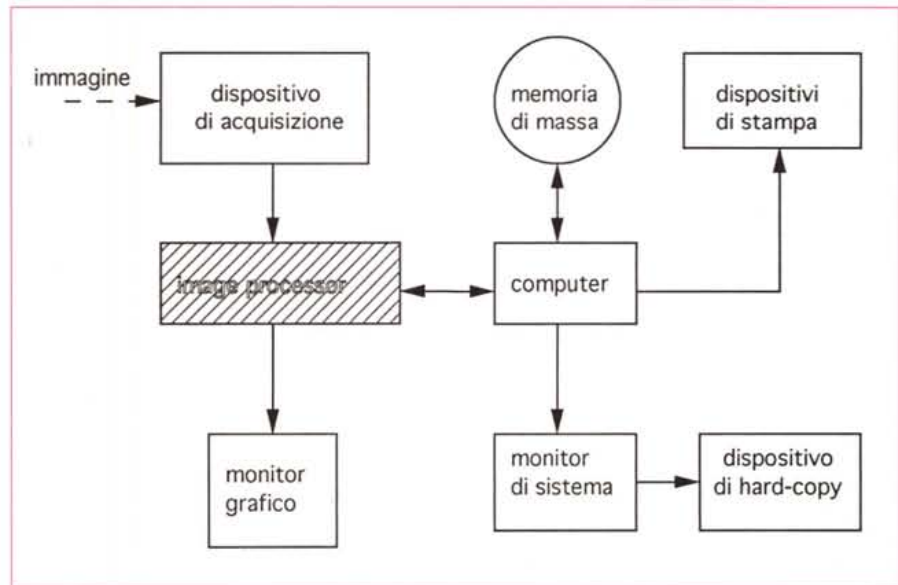


Figura 2 - Schema di semplice sistema per l'elaborazione di immagini.

ne, a dire il vero il nostro approccio prescinde dal tipo di dispositivo usato, però possiamo perlomeno individuare due tipi di periferiche diversi: lo scanner e il frame grabber.

Lo scanner dispositivo ormai sufficientemente diffuso, nasce per acquisire fotocopie o disegni o qualsiasi altra immagine riprodotta su supporto cartaceo, in realtà, proprio come una fotocopiatrice, può restituire un'immagine di tutto ciò che è posto sul piano di acquisizione. L'acquisizione vera e propria è affidata in genere ad una barra di elementi fotosensitivi che tramite un dispositivo meccanico viene fatta scorrere lungo tutto il piano. Il numero di punti campionati è elevato, ciò tuttavia comporta lunghi tempi di acquisizione, nell'ordine dei secondi, la riproduzione per contro è molto fedele. La risoluzione tipica va dai 75 ai 400 punti per pollice e perciò una pagina formato A4 di dimensioni $27,9 \times 29,7$ occupa dai 400 kb ai 2 Mb acquisendo a 256 livelli di grigio.

Il frame grabber invece è progettato per acquisire un segnale di tipo televisivo, per esempio generato da una telecamera, la sua caratteristica peculiare

è il campionamento delle immagini con un basso numero di punti, né avrebbe senso acquisire un numero di punti più elevato di quanti ne fornisca il segnale stesso. La tecnologia permette in questi casi tempi di acquisizione molto elevati, nel senso che sono ormai comuni frame grabber che acquisiscono l'immagine televisiva a frequenza video, permettendo in tal modo la ripresa di scene in movimento. In questo caso la risoluzione non è più misurata in punti per pollice ma in pixel ed in genere oscilla tra i 256×256 e i 1024×1024 pixel con un'occupazione di memoria dai 64 kb ad 1 Mb.

I frame grabber che si trovano in commercio però sono soltanto una parte di un sistema più complesso che si occupa di mettere a disposizione dell'utente hardware specializzato per l'elaborazione di immagini, permettendo così funzioni avanzate ed eseguite in modo veloce. Se tale dispositivo è aggiunto come periferica di un computer general-purpose, disporrà in generale di una propria uscita monitor.

Nel caso di acquisizione da scanner o se il frame grabber non consente elaborazioni locali, l'immagine deve essere trasferita nella memoria di un computer per eseguire le eventuali elaborazioni. Tale trasferimento, a meno che il sistema non preveda un modo preferenziale e specializzato, è inevitabilmente lento dovendo passare per il bus di sistema.

In figura 2 sono evidenziati anche le periferiche di memoria di massa e i dispositivi di output. Le prime meritano

$$f(x,y) = \begin{bmatrix} f(0,0) & \dots & f(0,n-1) \\ \dots & \dots & \dots \\ f(n-1,0) & \dots & f(n-1,n-1) \end{bmatrix}$$

Figura 3 - Rappresentazione in formule di un'immagine digitalizzata. È in effetti una matrice $N \times N$.

un posto di rilievo per il fatto che le immagini occupano molto spazio ed è quindi necessario che per applicazioni intensive si predispongano dispositivi molto capaci; per mantenere archivi on-line è ormai uso utilizzare periferiche a tecnologia ottica o magneto-ottica.

Per quanto riguarda i dispositivi di output, particolare attenzione bisogna porre al monitor che viene utilizzato per la visualizzazione, in quanto ci restituirà direttamente il risultato delle nostre elaborazioni. Premesso che la risoluzione di un monitor non può essere paragonata a quella ottenibile da una stampa su carta ad esempio, un buon monitor può permettere risoluzioni di 100 punti per pollice circa.

Il trasferimento su supporto cartaceo è affidato alle stampanti nelle diverse tecnologie oggi a disposizione, ma più che soffermarci su questo aspetto, è essenziale notare che la natura dell'immagine determina la qualità della stampa. Se ho acquisito a 400 dpi (dot per inch) allora ho informazioni necessarie a stampare a tale risoluzione e l'immagine risulterà di elevata qualità. Nel caso in cui abbia acquisito a risoluzione video, potrò al massimo permettermi un hard-copy dello schermo con una qualità molto inferiore alla precedente.

Un modello fisico dell'immagine

Abbiamo già stabilito che un'immagine è una funzione bidimensionale $f(x,y)$ dell'intensità luminosa nel punto di coordinate x,y . Dato che la luce è una forma di energia, $f(x,y)$ deve essere positiva e finita, in formule

$$0 < f(x,y) < \infty$$

Le immagini che ognuno di noi percepisce ogni giorno consistono di luce riflessa dagli oggetti. La funzione $f(x,y)$ può perciò essere considerata come caratterizzata da due componenti. La prima componente è la quantità di luce della sorgente incidente sulla scena, l'altra componente deriva dalla quantità di luce riflessa dagli oggetti della scena. Queste componenti sono chiamate componenti di illuminazione e di riflettanza e le indicheremo con $i(x,y)$ e $r(x,y)$ rispettivamente. Come vedete entrambe sono funzione di punto, ed in effetti si potrebbe obiettare che la componente di illuminazione sia invece una funzione dell'intera scena e non di un punto. Nella pratica spesso succede così, ma non si può escludere che ci siano zone dell'immagine più o meno illuminate a causa di ostacoli ad esempio. Il prodotto di queste due nuove funzioni restituisce la funzione $f(x,y)$:

$$f(x,y) = i(x,y) * r(x,y)$$

dove

$$0 < i(x,y) < \infty$$

e

$$0 < r(x,y) < 1$$

Quest'ultima disequazione indica che la riflettanza assume valore 0 quando l'oggetto assorbe totalmente la luce incidente mentre vale 1 quando c'è riflessione totale ed è determinata soltanto dalle caratteristiche dell'oggetto stesso.

I limiti delle precedenti equazioni sono ovviamente teorici considerando che in una giornata di pieno sole l'illuminazione può raggiungere i 10000 lux mentre con sole coperto si aggira intorno ai 1000 lux, una serata di luna piena offre circa 0.01 lux di illuminazione. In un interno l'illuminazione si stabilizza intorno ai 100 lux. Similmente valori tipici per $r(x,y)$ sono: 0.01 per il velluto nero, 0.65 per l'acciaio, 0.80 per una parete piana bianca, 0.90 per l'argento lucidato e 0.93 per la neve.

Considerati tali vincoli la funzione $f(x,y)$ assumerà valori tra 0.001 e 9300; se, come abbiamo fatto finora, consideriamo immagini a livelli di grigio, i due valori così calcolati costituiranno gli estremi della scala dei grigi che utilizzeremo. In pratica l'intervallo è spostato tra [0, L] dove a 0 è associato il nero e a L il bianco. Tutti i valori intermedi sono le tonalità dei grigi dal nero al bianco.

Campionamento e quantizzazione

Per poter eseguire una qualsiasi elaborazione elettronica tramite computer è necessario che la nostra funzione dell'immagine $f(x,y)$ sia digitalizzata spazialmente che in ampiezza. La digitalizzazione delle coordinate spaziali (x,y) viene indicata come campionamento immagine, mentre la digitalizzazione dell'ampiezza è chiamata quantizzazione del livello di grigio.

Supponiamo, senza perdita di generalità, che un'immagine continua $f(x,y)$ sia approssimata con campioni arrangiati secondo una matrice $N \times N$ come mostrato in figura 3, dove ciascun elemento è un numero intero. La matrice in figura è l'immagine digitale e ciascun elemento è un pixel.

Il valore N scelto come numero dei campioni è un parametro assai significativo in quanto determina la qualità dell'immagine come è ovvio; è pratica comune scegliere $N=2^n$ per ragioni di semplicità e il numero dei livelli di grigio

$G=2^n$. Inoltre abbiamo assunto che non ci sia nessuna ragione perché l'intervallo [0, L] non venga distribuito uniformemente sui G livelli.

L'occupazione di memoria dell'immagine sarà dunque $b=N \times N \times m$. Per ragioni pratiche però se m è fissato al valore 5 si preferisce comunque assegnare un byte intero per pixel, perché le operazioni di impaccamento dei bit sono assai onerose dal punto di vista computazionale.

Dato che l'immagine digitale è un'approssimazione dell'immagine continua, è ragionevole chiedersi quanti campioni e quanti livelli di grigio siano necessari per una buona approssimazione.

La risoluzione è chiaramente dipendente sia da N che da m , aumentando questi valori l'immagine digitale approssimerà sempre più fedelmente l'immagine continua. Chiaramente ci sono dei limiti fisici per cui non è possibile incrementare indefinitamente questi limiti.

Come ci si potrebbe aspettare è assai difficile definire in assoluto una «buona» immagine, in quanto i requisiti di qualità dipendono fortemente dall'applicazione. Vediamo allora che effetto può avere variare la griglia di campionamento spaziale dell'immagine. In figura 4 da a) ad f) notate la stessa immagine con $N=512, 256, 128, 64, 32, 16$ rispettivamente, mentre il numero dei livelli di grigio rimane immutato. Potete notare come la qualità decresca rapidamente al diminuire del numero di campioni e sopraggiunga l'effetto scacchiera.

In figura 5 da a) ad h), invece possiamo osservare la variazione del numero dei livelli di grigio su un'immagine dove N rimane costante pari a 512 mentre m varia da 8 a 1.

È anche ovvio che il numero di campioni e il numero dei livelli di grigio dipende in qualche misura dalla natura stessa dell'immagine che viene riprodotta; come riferimento, i requisiti necessari per ottenere immagini di qualità comparabile a quelle della TV in bianco e nero, possono fissarsi abbastanza ragionevolmente in 512×512 con 128 livelli di grigio. Un sistema minimo non dovrebbe mai scendere sotto i 256×256 con 64 livelli di grigio.

Una soluzione sofisticata per aumentare la risoluzione senza penalizzare l'occupazione di memoria consiste nell'effettuare un campionamento non uniforme, cioè campionare con maggior accuratezza le zone con maggior dettaglio o di maggior interesse e diminuire il numero di campioni nel resto dell'immagine. Questo modo di digitalizzare l'immagine però presuppone l'identificazione dei bordi della zona da campionare con maggior frequenza ed in gene-

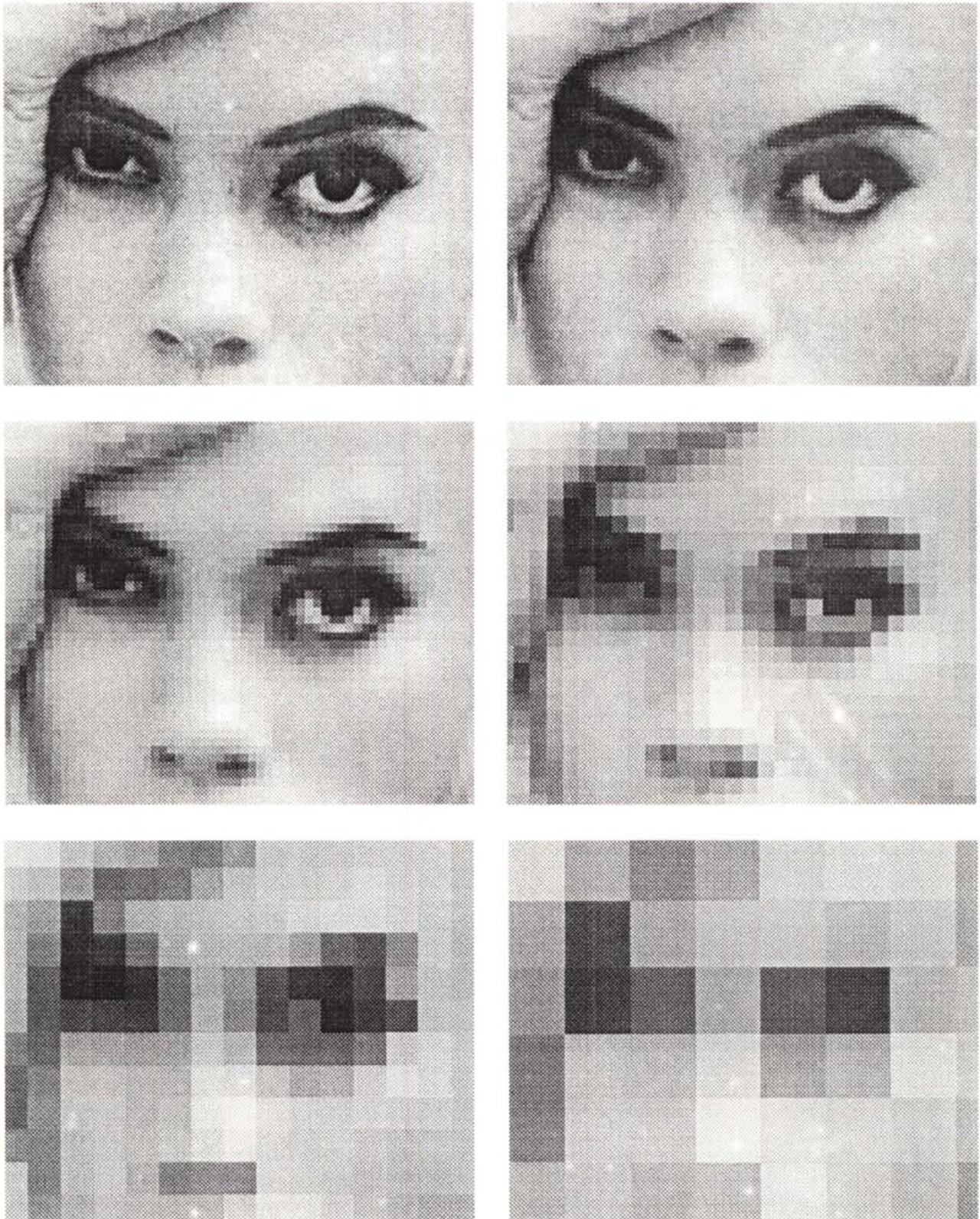


Figura 4 - In questa successione sono mostrati i cambiamenti di risoluzione da 512x512 a 16x16. I livelli di grigio sono mantenuti pari a 256.

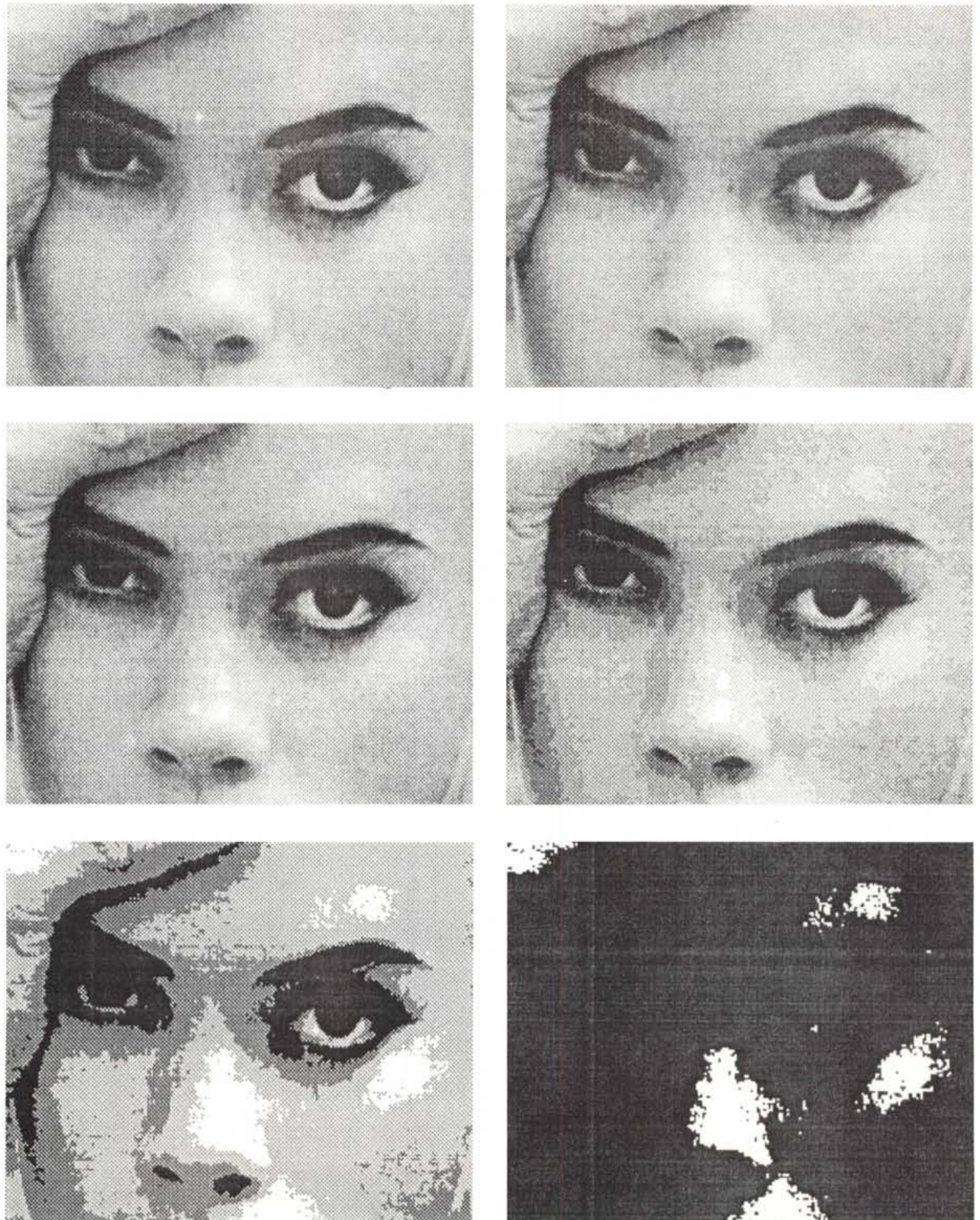
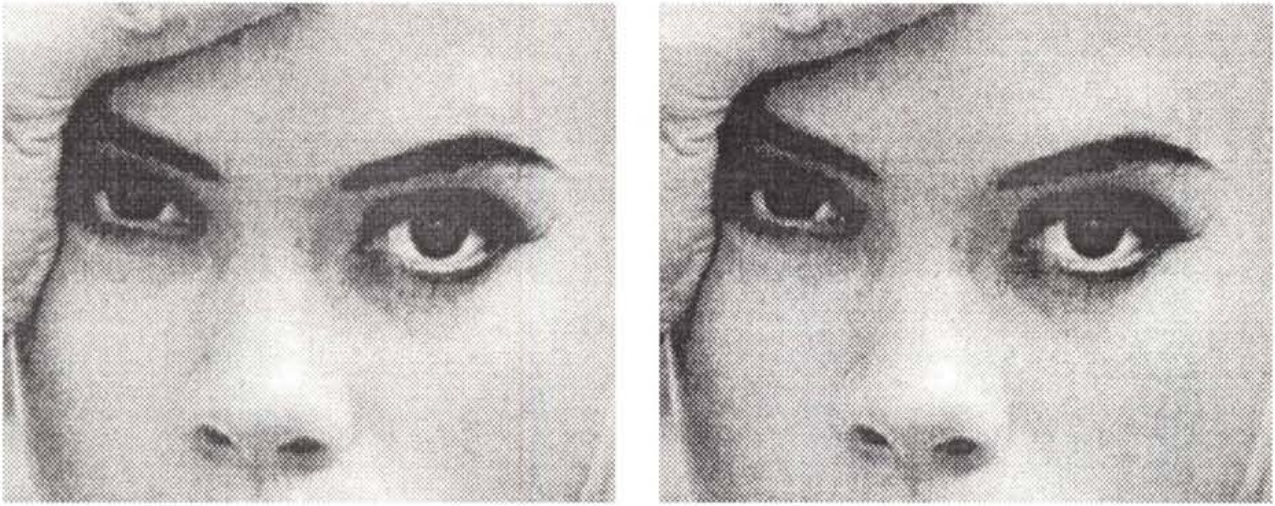


Figura 5 - Variazione dei livelli di grigio da 256 a 2. La qualità degrada rapidamente dopo 64 livelli.



rale non è un compito semplice.

Lo stesso discorso può essere fatto sui livelli di grigio, infatti è possibile utilizzare un numero di livelli di grigio per una zona di maggior interesse mentre per le restanti se ne utilizza un numero assai più basso; tuttavia anche questa quantizzazione presuppone l'individuazione della zona.

Una tecnica più attraente è invece basata sul calcolo della frequenza di occorrenza dei livelli di grigio; si utilizzerà una quantizzazione più fine nell'intervallo dove sono addensati la maggior parte dei grigi mentre se ne adopererà una più larga dove le frequenze di occorrenza sono minori.

Relazione di base fra i pixel; il concetto di Adiacenza fra pixel

Puntualizzeremo ora qualche fondamentale primitiva indispensabile per proseguire nel nostro discorso, sono semplici ma importanti relazioni fra i pixel di un'immagine digitale.

La prima relazione fondamentale è la definizione degli Adiacenti di un pixel. Fissato un pixel p di coordinate (x,y) , sono individuati 4 pixel adiacenti in orizzontale ed in verticale di coordinate:

$(x+1,y)$, $(x-1,y)$, $(x,y+1)$, $(x,y-1)$

Questi 4 pixel sono tutti a distanza unitaria da (x,y) e se p è un pixel di bordo alcuni dei suoi 4 adiacenti saranno esterni all'immagine digitale.

I 4 pixel adiacenti a p in diagonale sono invece individuati dalle coordinate:

$(x+1,y+1)$, $(x+1,y-1)$, $(x-1,y+1)$, $(x-1,y-1)$

Questi ultimi 4 punti insieme con i 4 precedenti costituiscono l'insieme degli 8-Adiacenti di p e sono un insieme molto importante, come vedremo nei prossimi appuntamenti.

La Connessione fra pixel

La Connessione fra pixel è un importante concetto utilizzato per individuare i confini delle zone che in qualche modo presentano delle caratteristiche di coerenza in un'immagine. Per stabilire se due pixel sono connessi dobbiamo determinare se sono adiacenti in qualche modo e se i loro livelli di grigio soddisfano uno specifico criterio di similarità, ad esempio se sono uguali.

Per fissare meglio le idee supponia-

mo che un criterio di connessione sia che i livelli di grigio abbiano un valore compreso tra 59 e 61, allora come in figura 6 si tratta di determinare un cammino tra i pixel in modo da individuare il confine di una regione. È chiaro che verrà verificata la coerenza soltanto per i pixel che sono adiacenti, nel senso che un cammino è definito soltanto se ci si sposta fra pixel adiacenti; di conseguenza è possibile definire cammini diversi se si permetterà adiacenza orizzontale e verticale o anche fra i 4 pixel diagonali. In formule possiamo definire un cammino da un pixel p di coordinate (x_0,y_0) ad un pixel q individuato dalle coordinate (s,t) come una sequenza di pixel distinti di coordinate

(x_0,y_0) , (x_1,y_1) ,....., (x_n,y_n) , (s,t)

in cui il generico pixel (x_i,y_i) è adiacente al pixel (x_{i-1},y_{i-1}) , con $1 \leq i \leq n$ dove n è la lunghezza del cammino. L'adiacenza è definita invece nel paragrafo precedente.

Conclusioni

L'elaborazione digitale delle immagini è una materia assai affascinante che spinge alla creatività e permette applicazioni molto varie e diverse fra loro. Nella nostra trattazione abbiamo intenzione di spiegare i principi in base ai quali possono eseguirsi le trasformazioni sulle immagini per gli scopi più particolari. Di volta in volta, mostreremo la teoria e le applicazioni più importanti e soddisferemo le curiosità che i lettori che ci seguono vorranno sottoporci.

MB

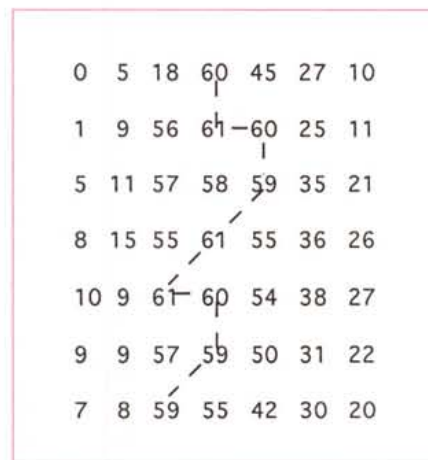


Figura 6 - Esempificazione di cammino su un'immagine. Come potete vedere vengono connessi i pixel adiacenti con valori compresi tra {59 e 61}.