

Edge preserving denoising in radiografia digitale

I. Frosio, M. Lucchese, N. A. Borghese

Statistica del rumore sull'immagine radiografica

Sia $\{g_{n,i}\}_{i=1..N}$ un'immagine radiografica digitale composta da N pixels, dove $g_{n,i}$ indica il livello di grigio assunto l'i-esimo pixel. Tale immagine sarà corrotta da rumore.

Come noto, il rumore preponderante in radiografia è associato con il processo di conteggio dei fotoni nello spettro dei raggi X, in arrivo sul sensore. In particolare, la statistica di conteggio dei fotoni è di tipo poissoniano. Diamo quindi di seguito una breve descrizione della distribuzione di Poisson.

Sia p_i il numero medio di fotoni contati dall'i-esimo pixel (si pensi di iterare un gran numero di volte l'esperimento in cui una sorgente emette idealmente lo stesso numero di fotoni, che vengono contati dall'i-esimo pixel del sensore). Possiamo pensare a p_i come il numero "corretto" (non rumoroso) di fotoni associati con il pixel i-esimo. La possibilità di misurare un numero $p_{n,i}$ di fotoni è allora data da:

$$p_{\text{Pois}}(p_{n,i} | p_i) = \frac{p_i^{p_{n,i}} e^{-p_i}}{p_{n,i}!} \quad (1).$$

Dal momento che il sensore restituisce un livello di grigio e non un numero di fotoni, è necessario considerare un modello per il sensore che descriva la trasformazione da fotoni a livello di grigio. Si noti che i fotoni in arrivo sulle singole cellette del sensore sono già affetti da rumore poissoniano. Tali fotoni impattano contro un materiale scintillatore, che converte i fotoni nello spettro dei raggi X in fotoni visibili con una data efficienza. Tale processo potrebbe introdurre ulteriore rumore, ma trascuriamo questo aspetto. I fotoni visibili raggiungono quindi la celletta di un CCD, dove vengono trasformati in elettroni. Anche in questo caso entrano in gioco un'efficienza caratteristica ed una potenziale introduzione di nuovo rumore; anche in questo caso trascuriamo questo rumore. Infine, il CCD viene "letto" da un ADC che trasforma il voltaggio associato con gli elettroni in un livello di grigio. Anche l'ADC potrebbe introdurre rumore. E' comunque lecito trascurare il rumore introdotto dallo scintillatore, dal CCD e dall'ADC coverter in quanto, per un sistema ben assemblato, è la statistica del conteggio dei fotoni a determinare in modo significativo l'SNR sull'immagine a livello di grigio. In toto, la funzione di trasferimento del sensore può essere descritta nel modo seguente:

$$g_{n,i} = G \cdot p_{n,i} + O \quad (2),$$

dove G ed O rappresentano rispettivamente il gain e l'offset del sensore, e sono assunti costanti per l'intero sensore. Sotto l'ipotesi che il sensore sia opportunamente calibrato, il sensore restituisce un livello di grigio 0 quando nessun fotone raggiunge la sua superficie. Considerando quindi l'uscita di un sensore calibrato possiamo trascurare il termine di offset ($O=0$).

Dalla (1) e dalla (2) è allora possibile derivare la densità di probabilità per la variabile casuale $g_{n,i}$ (livello di grigio rumoroso del pixel i-esimo), dato il valore non rumoroso per il pixel i-esimo, g_i , e sotto l'ipotesi che l'offset del sensore sia nullo. Tale densità di probabilità risulta essere uguale a:

$$p_{\text{PC}}(g_{n,i} | g_i) = \frac{1}{G} \cdot \left[\left(\frac{g_i}{G} \right)^{\frac{g_{n,i}}{G}} e^{-\frac{g_i}{G}} \right] / \left[\left(\frac{g_{n,i}}{G} \right)! \right] \quad (3).$$

La (3) mette in risalto che affermare che la statistica del rumore in radiografia (sui livello di grigio dell'immagine) è Poissiniana è un'affermazione non completamente corretta, in quanto nell'espressione della densità di probabilità compare il gain del sensore. Solamente nel caso di gain unitario possiamo affermare che il rumore sui livello di grigio di un'immagine è poissoniano.

Definizione della funzione di verosimiglianza

Per effettuare il denoising di un'immagine radiografica affetta da rumore fotonico, possiamo considerare la funzione di verosimiglianza. Tale funzione descrive la probabilità di misurare i valori di grigio $\{g_{n,i}\}_{i=1..N}$, dati i livelli di grigio non rumorosi $\{g_i\}_{i=1..N}$. Dal momento che i valori g_i non sono noti, la verosimiglianza è una funzione di $\{g_i\}_{i=1..N}$. Massimizzando la funzione di verosimiglianza rispetto a $\{g_i\}_{i=1..N}$, è possibile trovare i livelli di grigio $\{g_i\}_{i=1..N}$ tali per cui è massima la probabilità di misurare i dati rumorosi $\{g_{n,i}\}_{i=1..N}$.

Per scrivere tale funzione, facciamo l'ipotesi che i valori registrati in due pixel diversi (e quindi anche il rumore) siano indipendenti l'uno dall'altro. Tale ipotesi equivale a considerare un sensore "perfetto", la cui risposta all'impulso sia la delta di Dirac. Sotto tale ipotesi, la funzione di verosimiglianza è data dal prodotto delle probabilità per ciascun pixel misurato, quindi:

$$L(g_{n,i} | G, g_i) = \prod_{i=1}^N p_{PC}(g_{n,i} | g_i) \quad (4).$$

Invece di massimizzare (4), che contiene N prodotti, risulta più facile minimizzare il logaritmo negativo della verosimiglianza, che consiste in una somma di logaritmi. Infatti:

$$f(g_{n,i} | G, g_i) = -\ln[L(g_{n,i} | G, g_i)] = -\ln\left[\prod_{i=1}^N p_{PC}(g_{n,i} | g_i)\right] = -\sum_{i=1}^N \ln[p_{PC}(g_{n,i} | g_i)] \quad (5),$$

dove f è la funzione che deve essere minimizzata per massimizzare la verosimiglianza. Scriviamo per esteso f :

$$\begin{aligned} f(g_{n,i} | G, g_i) &= -\sum_{i=1}^N \ln \left[\frac{1}{G} \cdot \left[\left(\frac{g_i}{G} \right)^{\frac{g_{n,i}}{G}} e^{-\frac{g_i}{G}} \right] / \left[\left(\frac{g_{n,i}}{G} \right)! \right] \right] = -\sum_{i=1}^N \left\{ -\ln(G) + \frac{g_{n,i}}{G} \ln \left(\frac{g_i}{G} \right) - \frac{g_i}{G} - \ln \left[\left(\frac{g_{n,i}}{G} \right)! \right] \right\} = \\ &= -\sum_{i=1}^N \left\{ -\ln(G) + \frac{g_{n,i}}{G} \ln(g_i) - \frac{g_{n,i}}{G} \ln(G) - \frac{g_i}{G} - \ln \left[\left(\frac{g_{n,i}}{G} \right)! \right] \right\} = \\ &= -\sum_{i=1}^N \left\{ -\ln(G) - \frac{g_{n,i}}{G} \ln(G) - \ln \left[\left(\frac{g_{n,i}}{G} \right)! \right] \right\} - \frac{1}{G} \sum_{i=1}^N \{ g_{n,i} \ln(g_i) - g_i \} \end{aligned} \quad (6).$$

Dalla (6) è evidente che la f è scomponibile in due termini. Il primo termine non dipende da g_i , quindi è costante una volta che sia data l'immagine misurata $\{g_{n,i}\}_{i=1..N}$; tale termine non entrerà dunque nel processo di minimizzazione di f . Il secondo termine è invece il classico termine utilizzato in tomografia ad emissione (dove la natura poissoniana del rumore è evidente) per descrivere la distanza tra un'immagine rumorosa $\{g_{n,i}\}_{i=1..N}$ e un'immagine non rumorosa $\{g_i\}_{i=1..N}$. A parte una costante ed il fattore di scala $1/G$, tale termine coincide con la distanza di Kulback-Leibler (KL) di $\{g_{n,i}\}_{i=1..N}$ e $\{g_i\}_{i=1..N}$. Dunque per la massimizzazione della verosimiglianza è necessario minimizzare la KL tra $\{g_{n,i}\}_{i=1..N}$ e $\{g_i\}_{i=1..N}$:

$$KL(g_{n,i}, g_i) = -\sum_{i=1}^N \left\{ g_{n,i} \ln \left(\frac{g_i}{g_{n,i}} \right) - g_i + g_{n,i} \right\} \quad (7a).$$

Tenendo conto che i termini $\ln(g_{n,i})$ e $g_{n,i}$ non sono funzione di g_i , si può minimizzare la:

$$\tilde{f}(g_{n,i} | G, g_i) = -\sum_{i=1}^N \{ g_{n,i} \ln(g_i) - g_i \} \quad (7b).$$

Per la minimizzazione della (7b), proviamo a porre le derivate di $\tilde{f}(g_{n,i} | G, g)$ rispetto a $\{g_i\}_{i=1..N}$ uguali a zero. Otteniamo in tal caso:

$$\frac{\partial \tilde{f}(g_{n,i} | G, g_i)}{\partial g_k} = \frac{\partial}{\partial g_k} \left[- \sum_{i=1}^N \{g_{n,i} \ln(g_i) - g_i\} \right] = \frac{\partial}{\partial g_k} \left[- \{g_{n,k} \ln(g_k) - g_k\} \right] = - \left(\frac{g_{n,k}}{g_k} - 1 \right) = 0 \Rightarrow$$

$$\Rightarrow \left(\frac{g_{n,k}}{g_k} - 1 \right) = 0 \Rightarrow g_k = g_{n,k}$$
(8)

Dunque la massimizzazione della verosimiglianza ci porta in questo caso ad una soluzione banale: l'immagine non rumorosa $\{g_i\}_{i=1..N}$ più probabile, data l'immagine rumorosa $\{g_{n,i}\}_{i=1..N}$, coincide esattamente con l'immagine rumorosa $\{g_{n,i}\}_{i=1..N}$.

Tale risultato, peraltro atteso, è un esempio estremo del problema di overfitting che spesso caratterizza gli approcci di stima alla massima verosimiglianza. In questo caso, non essendoci alcun legame tra un pixel ed i suoi vicini, la stima massimamente probabile coincide esattamente con l'immagine rumorosa registrata, quindi la soluzione trovata non è di alcun interesse pratico. E' però da notare che, in questa sezione, abbiamo definito la KL come misura di distanza tra un'immagine rumorosa e un'immagine non rumorosa. D'ora in poi chiameremo tale termine "termine di data fitting", in quanto esso quantifica la bontà del fitting tra la soluzione $\{g_i\}_{i=1..N}$ ed i dati misurati $\{g_{n,i}\}_{i=1..N}$.

Teoria della regolarizzazione

Il problema di overfitting della verosimiglianza può essere risolto ricorrendo alla cosiddetta teoria della regolarizzazione. Possiamo cioè pensare di stimare un'immagine non rumorosa $\{g_i\}_{i=1..N}$ che sia il più vicino possibile ai dati misurati $\{g_{n,i}\}_{i=1..N}$, dunque che minimizzi la KL, ma che allo stesso tempo non sia dotata di alcune proprietà indesiderate, descritte da un opportuno termine, detto di regolarizzazione. La soluzione sarà quindi un compromesso tra la minimizzazione del termine di data fitting e quello di regolarizzazione.

Nell'ambito della teoria della regolarizzazione, questo si traduce nella minimizzazione rispetto a $\{g_i\}_{i=1..N}$ del seguente funzionale composto $J(g_n | g)$:

$$\begin{cases} J(g_{n,i} | g_i) = J_{DF}(g_{n,i} | g_i) + \mu \cdot J_R(g_i) \\ J(g_n | g) = \sum_{i=1}^N J(g_{n,i} | g_i) \end{cases} \quad (9),$$

dove J_{DF} e J_R rappresentano rispettivamente il termine di data fitting (la KL per rumore di conteggio dei fotoni, poissoniano) e il termine di regolarizzazione. Si noti che il termine di regolarizzazione è una funzione della sola immagine $\{g_i\}_{i=1..N}$, in quanto esso descrive in maniera quantitativa una qualità dell'immagine soluzione che vogliamo penalizzare. Vi è inoltre un termine scalare μ , che pesa il termine di regolarizzazione; per $\mu=0$, minimizzando J si ottiene la soluzione alla massima verosimiglianza; al crescere del valore di μ , minimizzando J si ottiene una soluzione sempre più distante dai dati misurati, ma più regolarizzata.

Una scelta classica per il termine di regolarizzazione è la norma del gradiente dell'immagine. Dal momento che il rumore genera frequentemente delle oscillazioni locali nel segnale (il rumore è in genere ad alta frequenza), come conseguenza esso genera anche un incremento della norma del gradiente. Dunque la minimizzazione della norma del gradiente permette di stimare un'immagine in cui sia minimizzato il contributo del rumore. Ciò permette quindi di evitare il problema di overfitting dell'approccio alla massima verosimiglianza.

Per riassumere, viene minimizzata una cifra di merito data da due termini: il primo (termine di data fitting) obbliga a cercare un'immagine il più vicina possibile ai dati misurati; il secondo (termine di regolarizzazione) penalizza le immagini troppo rumorose. Il parametro μ bilancia il peso di ciascuno dei due termini, e viene scelto in maniera empirica dall'utente.

Per ragioni storiche, in qualche modo legate anche alla semplicità di calcolo, uno dei primi termini proposti per la regolarizzazione è la norma del gradiente al quadrato:

$$J_R(g_i) = \|\nabla g_i\|_2^2 \quad (10)$$

Si parla in questo caso di regolarizzazione alla Tikonov. Come vedremo più avanti, questo tipo di regolarizzazione corrisponde a fare l'ipotesi che la norma del gradiente nell'immagine abbia una distribuzione di tipo gaussiano a media nulla. La formulazione risulta essere semplice dal punto di vista computazionale grazie all'elevamento a quadrato della norma del gradiente. In questo caso, infatti, il termine di regolarizzazione può essere scritto come la somma di due quadrati ($g_x^2 + g_y^2$, dove g_x e g_y indicano rispettivamente il gradiente in direzione x e y); tale termine, oltre che facilmente calcolabile, è anche facilmente differenziabile (la differenziazione è necessaria per effettuare l'ottimizzazione di J con un metodo iterativo di ordine uguale o superiore al primo).

Tuttavia la regolarizzazione alla Tikonov ha un grosso difetto. Le immagini reali sono spesso caratterizzate dalla presenza di diversi oggetti, ciascuno caratterizzato da un livello di grigio abbastanza omogeneo. In tali immagini riconosciamo due tipi di gradiente non nullo: vi sono gradienti non nulli, in genere di piccola norma, dovuti alle oscillazioni ad alta frequenza del rumore; e gradienti non nulli di grande norma, associati con le transizioni tra i diversi oggetti (edge). Un algoritmo di denoising "intelligente" dovrebbe essere in grado di eliminare le oscillazioni dovute al rumore, preservando allo stesso tempo gli edge dell'immagine; l'approccio alla Tikonov non è però in grado di garantire tali prestazioni. Il motivo di tale "cattivo" comportamento sta tutto nell'esponente 2 scelto per la norma del gradiente. Il termine di penalizzazione di Tikonov tende a sfavorire in maniera eccessiva gli edge in cui la norma del gradiente è troppo grande (Fig. 1), dal momento che la penalizzazione va con il quadrato della norma del gradiente. Ciò che si ottiene come risultato è un effetto di smoothing dell'immagine, nella quale il rumore viene fortemente ridotto, ma allo stesso tempo vengono distrutti anche gli edge significativi.

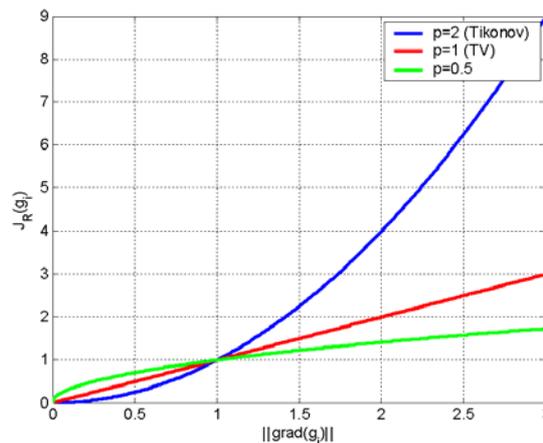


Fig. 1. Andamento del termine di regolarizzazione in funzione della norma del gradiente, al variare dell'esponente p .

Risultati migliori possono essere ottenuti variando il valore dell'esponente per la norma del gradiente. In questo modo, si può modulare il modo in cui viene penalizzata la norma del gradiente, cercando di ovviare al problema della sovrappenalizzazione degli edge, tipico dell'approccio alla Tikonov. In generale, possiamo dunque assumere per il termine di regolarizzazione la seguente funzione:

$$J_R(g_i) = \|\nabla g_i\|_2^p \quad (11),$$

dove il termine p è uno scalare positivo.

Per $p=2$ si ottiene come caso particolare il termine di "regolarizzazione alla Tikonov".

Per $p=1$ si ottiene una crescita lineare del termine di regolarizzazione al crescere della norma del gradiente e l'approccio prende il nome di regolarizzazione "total variation" (TV).. E' evidente che in questo caso gli edge dell'immagine non saranno penalizzati in maniera "pesante", così come accade con Tikonov; l'approccio "Total variation" è universalmente riconosciuto all'interno della comunità scientifica come un approccio utile per effettuare un "edge preserving denoising". Tale approccio sarà quello adottato in questo lavoro.

Per $p < 1$, si ottengono altri termini di penalizzazione validi per la creazione di algoritmi per l'“edge preserving denoising”. Tuttavia tali esponenti non hanno trovato vasta applicazione in letteratura, e quindi non saranno considerati ulteriormente (a parte per quanto riguarda la possibilità di effettuare regolarizzazione alla Geman & Geman, si veda più avanti). Il motivo di ciò è probabilmente il fatto che J_R non è convesso per $p < 1$ e quindi risulta più difficile da minimizzare.

Per concludere, è possibile risolvere il problema di overfitting della verosimiglianza introducendo un termine di regolarizzazione nella cifra di merito da minimizzare. Allo scopo di sviluppare un algoritmo “edge preservino denoising”, la scelta cade su un termine di regolarizzazione tipo “norma del gradiente – TV”.

Interpretazione Bayesiana della teoria della regolarizzazione (stima MAP)

Il termine di regolarizzazione, introdotto in modo “empirico” al paragrafo precedente per risolvere il problema di overfitting dell'approccio alla massima verosimiglianza, ha in realtà un'interpretazione più profonda dal punto di vista bayesiano. In particolare, è possibile dimostrare che tale termine rappresenta l'aderenza della soluzione rispetto ad un'informazione data a priori sulle proprietà della soluzione stessa. Daremo di seguito una dimostrazione esplicita di tale affermazione per il caso di regolarizzazione alla Tikonov. In fondo al paragrafo, generalizzeremo la conclusione ad altri termini di regolarizzazione.

Si supponga dunque di volere stimare i livelli di grigio non rumorosi di una radiografia $\{g_i\}_{i=1..N}$, data la radiografia misurata sul campo $\{g_{n,i}\}_{i=1..N}$. Si faccia inoltre l'ipotesi che i gradienti x e y nell'immagine abbiano entrambi una distribuzione gaussiana a media nulla e varianza σ , cioè:

$$\begin{cases} p(g_{x,i}) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma}} \cdot e^{-\frac{1}{2} \left(\frac{g_{x,i}}{\sigma} \right)^2} \\ p(g_{y,i}) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma}} \cdot e^{-\frac{1}{2} \left(\frac{g_{y,i}}{\sigma} \right)^2} \end{cases} \quad (12),$$

dove $g_{x,i}$ e $g_{y,i}$ indicano rispettivamente il gradiente in direzione x e y dell'immagine. Se il gradiente dell'immagine è indipendente dal livello di grigio (ipotesi peraltro più che plausibile), la probabilità di misurare il livello di grigio $g_{n,i}$ e i gradienti $g_{x,i}$ e $g_{y,i}$ nel pixel i -esimo è data dal prodotto delle tre probabilità corrispondenti, cioè:

$$\begin{aligned} p(g_i, g_{x,i}, g_{y,i}) &= p(g_i) \cdot p(g_{x,i}) \cdot p(g_{y,i}) = \\ &= p_{PC}(g_{n,i} | g_i) \cdot \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma}} \cdot e^{-\frac{1}{2} \left(\frac{g_{x,i}}{\sigma} \right)^2} \cdot \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma}} \cdot e^{-\frac{1}{2} \left(\frac{g_{y,i}}{\sigma} \right)^2} \end{aligned} \quad (13).$$

Proviamo allora a massimizzare la probabilità di misurare l'immagine rumorosa $\{g_{n,i}\}_{i=1..N}$ rispetto all'immagine (incognita) non rumorosa $\{g_i\}_{i=1..N}$, date le nostre ipotesi sulla distribuzione statistica del gradiente nell'immagine. Per fare ciò dovremmo massimizzare il prodotto delle probabilità per ogni singolo pixel, ma risulta essere più facile minimizzare il logaritmo negativo di tale quantità. Vogliamo cioè minimizzare:

$$\begin{aligned}
 \hat{f}(g_n | G, g, \nabla g) &= -\sum_{i=1}^N \ln \left[p_{PC}(g_{n,i} | g_i) \cdot \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma}} \cdot e^{-\frac{1}{2} \left(\frac{g_{x,i}}{\sigma} \right)^2} \cdot \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma}} \cdot e^{-\frac{1}{2} \left(\frac{g_{y,i}}{\sigma} \right)^2} \right] = \\
 &= -\sum_{i=1}^N \left\{ \ln [p_{PC}(g_{n,i} | g_i)] + \ln \left[\frac{1}{\sqrt{2 \cdot \pi \cdot \sigma}} \cdot e^{-\frac{1}{2} \left(\frac{g_{x,i}}{\sigma} \right)^2} \right] + \ln \left[\frac{1}{\sqrt{2 \cdot \pi \cdot \sigma}} \cdot e^{-\frac{1}{2} \left(\frac{g_{y,i}}{\sigma} \right)^2} \right] \right\} = \\
 &= -\sum_{i=1}^N \left\{ -\ln(G) - \frac{g_{n,i}}{G} \ln(G) - \ln \left[\left(\frac{g_{n,i}}{G} \right)! \right] \right\} - \frac{1}{G} \sum_{i=1}^N \{g_{n,i} \ln(g_i) - g_i\} - \sum_{i=1}^N \left\{ 2 \cdot \ln \left[\frac{1}{\sqrt{2 \cdot \pi \cdot \sigma}} \right] - \frac{1}{2} \cdot \left(\frac{g_{x,i}}{\sigma} \right)^2 - \frac{1}{2} \cdot \left(\frac{g_{y,i}}{\sigma} \right)^2 \right\} = \\
 &= -\sum_{i=1}^N \left\{ -\ln(G) - \frac{g_{n,i}}{G} \ln(G) - \ln \left[\left(\frac{g_{n,i}}{G} \right)! \right] \right\} + 2 \cdot N \cdot \ln(\sqrt{2 \cdot \pi \cdot \sigma}) - \frac{1}{G} \sum_{i=1}^N \{g_{n,i} \ln(g_i) - g_i\} + \frac{1}{2 \cdot \sigma^2} \cdot \sum_{i=1}^N (g_{x,i}^2 + g_{y,i}^2) = \\
 &= \text{cost} - \frac{1}{G} \sum_{i=1}^N \{g_{n,i} \ln(g_i) - g_i\} + \frac{1}{2 \cdot \sigma^2} \cdot \sum_{i=1}^N (g_{x,i}^2 + g_{y,i}^2)
 \end{aligned} \tag{14}$$

nella (14), risulta evidente il contributo dato dall'ipotesi a priori sulla distribuzione gaussiana del gradiente, che corrisponde alla norma al quadrato equivalente ad una regolarizzazione tipo Tikonov (qde).

Il parametro di regolarizzazione, μ , è equivalente al rapporto $1/(2\sigma^2)$; è evidente che, nel caso in cui i gradienti dell'immagine siano molto piccoli ($\sigma \rightarrow 0$) è possibile applicare un parametro di regolarizzazione molto grande ($1/(2\sigma^2) \rightarrow \infty$), dal momento che è ragionevole penalizzare in maniera forte ogni gradiente presente nell'immagine. Al contrario, se l'immagine è caratterizzata dalla presenza di molti edge, la σ sarà grande ed il corrispondente parametro di regolarizzazione sarà piccolo. Tali considerazioni, estese ad esempio al caso di regolarizzazione TV, possono essere alla base di un metodo per il calcolo automatico del parametro di regolarizzazione migliore. La (14) dimostra dunque che, assumendo che i gradienti x e y di un'immagine abbiano una distribuzione gaussiana e siano indipendenti dal livello di grigio, è possibile derivare lo schema di regolarizzazione di Tikonov. Più in generale, possiamo una distribuzione di Gibbs per i gradienti nell'immagine:

$$p(g_{x,i}, g_{y,i}) = e^{-J_R(g_{x,i}, g_{y,i})} \tag{15}$$

una scelta opportuna della funzione $J_R(g_{x,i}, g_{y,i})$ porta, in modo analogo a quanto dimostrato sopra, alla formulazione di Tikonov, TV, etc

La (14) fornisce dunque la chiave per interpretare in senso statistico la teoria della regolarizzazione. Tale equazione può essere anche interpretata facendo ricorso al teorema di Bayes. Se scriviamo il teorema nel caso della distribuzione dei livelli di grigio e della norma dei gradienti di un'immagine, otteniamo:

$$p(\mathbf{g}_n | \mathbf{g}) \cdot p(\mathbf{g}) = p(\mathbf{g} | \mathbf{g}_n) \cdot p(\mathbf{g}_n) \Rightarrow p(\mathbf{g} | \mathbf{g}_n) = \frac{p(\mathbf{g}_n | \mathbf{g}) \cdot p(\mathbf{g})}{p(\mathbf{g}_n)} \tag{16}$$

Nella (16), bisogna tenere presente che la $p(\mathbf{g})$ descrive in realtà la probabilità di registrare un dato gradiente. La $p(\mathbf{g}_n | \mathbf{g})$ è la verosimiglianza (ovvero la probabilità di misurare i dati rumorosi \mathbf{g}_n data l'immagine reale \mathbf{g}) mentre la $p(\mathbf{g})$ rappresenta l'informazione a priori, ovvero la probabilità (a priori!) che l'immagine sia distribuita come \mathbf{g} , ovvero che il gradiente assuma una certa norma. La $p(\mathbf{g}_n)$ è costante, mentre la $p(\mathbf{g} | \mathbf{g}_n)$ è la probabilità a posteriori dell'immagine non rumorosa \mathbf{g} , data l'immagine rumorosa (misurata) \mathbf{g}_n . Da questo punto di vista, il problema di stima di \mathbf{g} diventa il problema di massimizzare la probabilità a posteriori di \mathbf{g} (si parla infatti di stima MAP, maximum a posteriori). In realtà poi, essendo $p(\mathbf{g}_n)$ solo una costante di normalizzazione che non dipende dalla soluzione \mathbf{g} , la stima MAP si effettua massimizzando:

$$MAP = p(\mathbf{g}_n | \mathbf{g}) \cdot p(\mathbf{g}) = \text{likelihood} \times \text{a priori} \tag{17}$$

ovvero minimizzando il suo logaritmo negativo.

Primi risultati sperimentali

I primi risultati sperimentali sono stati volti alla verifica della correttezza del modello scelto per la descrizione del rumore (ipotesi di rumore di conteggio dei fotoni preponderante) e delle potenzialità della teoria della regolarizzazione per l'edge preservino denoising in radiografia digitale.

In prima istanza si è quindi proceduto alla minimizzazione della (9), dove il termine di data fitting è costituito dalla KL, mentre per il termine di regolarizzazione si è proceduto in maniera empirica per valutare le prestazioni al variare dell'esponente p e del coefficiente μ . La minimizzazione viene effettuata mediante il metodo del gradiente, per il quale è necessario calcolare la derivata della cifra di merito rispetto ai pixel dell'immagine incognita, g_i . Per il calcolo della norma del gradiente sull'immagine, si considerino, tra gli otto vicini di ogni pixel, i soli due vicini in direzione nord e ovest (quindi $g_x(x,y)=g(x,y)-g(x-1,y)$ e $g_y(x,y)=g(x,y)-g(x,y-1)$). In tal caso si tenga presente che il termine incognito g_k compare: all'interno del termine di data fitting, per il solo elemento della sommatoria relativo all'elemento k -esimo; e all'interno del termine di regolarizzazione, negli elementi della sommatoria corrispondenti al gradiente calcolato nel pixel k -esimo e nei suoi due vicini in direzione est e sud. La cifra di merito da minimizzare e la sua derivata risultano dunque essere:

$$\begin{cases} J_{DF}(g_{n,i} | g_i) = -\sum_{i=1}^N \{g_{n,i} \ln(g_i) - g_i\} \\ J_R(g_i) = \sum_{i=1}^N \|\nabla g_i\|_2^p = \sum_{i=1}^N (g_{x,i}^2 + g_{y,i}^2)^{\frac{p}{2}} \\ J(g_{n,i} | g_i) = J_{DF}(g_{n,i} | g_i) + \mu \cdot J_R(g_i) \\ J(\mathbf{g}_n | \mathbf{g}) = \sum_{i=1}^N J(g_{n,i} | g_i) \end{cases}$$

$$\frac{\partial J(\mathbf{g}_n | \mathbf{g})}{\partial g_k} = -\sum_{i=1}^N \frac{\partial \{g_{n,i} \ln(g_i) - g_i\}}{\partial g_k} + \sum_{i=1}^N \frac{\partial \left\{ (g_{x,i}^2 + g_{y,i}^2)^{\frac{p}{2}} \right\}}{\partial g_k} = \quad (18)$$

$$= -\left(\frac{g_{n,k}}{g_k} - 1 \right) + \frac{\partial \left\{ \|\nabla g(u_k, v_k)\|_2^p + \|\nabla g(u_k + 1, v_k)\|_2^p + \|\nabla g(u_k, v_k + 1)\|_2^p \right\}}{\partial g(u_k, v_k)}$$

dove è stata usata la notazione (u_k, v_k) per indicare esplicitamente la posizione del k -esimo pixel. Continuando il calcolo della derivata si ottiene:

$$\begin{aligned} \frac{\partial J(\mathbf{g}_n | \mathbf{g})}{\partial g_k} &= -\frac{g_{n,k}}{g_k} + 1 + p \cdot \|\nabla g(u_k, v_k)\|_2^{p-1} \cdot \frac{\partial \left[\|\nabla g(u_k, v_k)\|_2 \right]}{\partial g(u_k, v_k)} + p \cdot \|\nabla g(u_k + 1, v_k)\|_2^{p-1} \cdot \frac{\partial \left[\|\nabla g(u_k + 1, v_k)\|_2 \right]}{\partial g(u_k, v_k)} + \\ &+ p \cdot \|\nabla g(u_k, v_k + 1)\|_2^{p-1} \cdot \frac{\partial \left[\|\nabla g(u_k, v_k + 1)\|_2 \right]}{\partial g(u_k, v_k)} = \\ &= -\frac{g_{n,k}}{g_k} + 1 + \\ &p \cdot \left\{ \|\nabla g(u_k, v_k)\|_2^{p-2} \cdot 2 \cdot [g_x(u_k, v_k) + g_y(u_k, v_k)] - \|\nabla g(u_k + 1, v_k)\|_2^{p-2} \cdot 2 \cdot g_x(u_k, v_k) - \|\nabla g(u_k, v_k + 1)\|_2^{p-2} \cdot 2 \cdot g_x(u_k, v_k + 1) \right\} \end{aligned} \quad (19)$$

la quale permette di calcolare effettuare la minimizzazione di J mediante il metodo del gradiente, ovvero di aggiornare iterativamente la soluzione $\{g_i\}_{i=1..N}$ secondo lo schema:

$$g_i^{(t+1)} = g_i^{(t)} + \alpha \cdot \frac{\partial J(\mathbf{g}_n | \mathbf{g}^{(t)})}{\partial g_k} \quad (20)$$

dove l'apice (t) indica l'iterazione, mentre α è un parametro scalare (chiamato damping) che viene ottimizzato ad ogni iterazione per garantire che J sia decrescente (in particolare il valore di α viene dimezzato durante la fase di line search nel caso in cui il valore corrente di α non garantisca la decrescita di J).

Si noti che nella (19), per $p < 2$, la norma del gradiente può comparire a denominatore; in tal caso, una norma del gradiente nulla genererebbe una divisione per zero; per evitare tale situazione, invece di considerare la norma del gradiente, si considera la seguente approssimazione:

$$\|\nabla g_i\|_2 = \sqrt{g_{x,i}^2 + g_{y,i}^2 + T^2} \quad (21),$$

dove T è uno scalare a scelta dell'utente sufficientemente piccolo da influenzare in modo poco significativo la norma del gradiente, ma sufficiente per evitare divisioni per zero (approssimazione di Hubert).

In Fig. 2 è mostrata un'immagine simulata, prima (a sinistra) e dopo (a destra) la corruzione dell'immagine con rumore determinato dal conteggio dei fotoni. In Fig. 3, 4 e 5 sono mostrati i risultati nel caso in cui si scelga $p=2$, $p=1$ e $p=0.5$, con l'errore misurato come differenza dell'immagine trattata con il filtro di denoising e l'immagine originale priva di rumore.

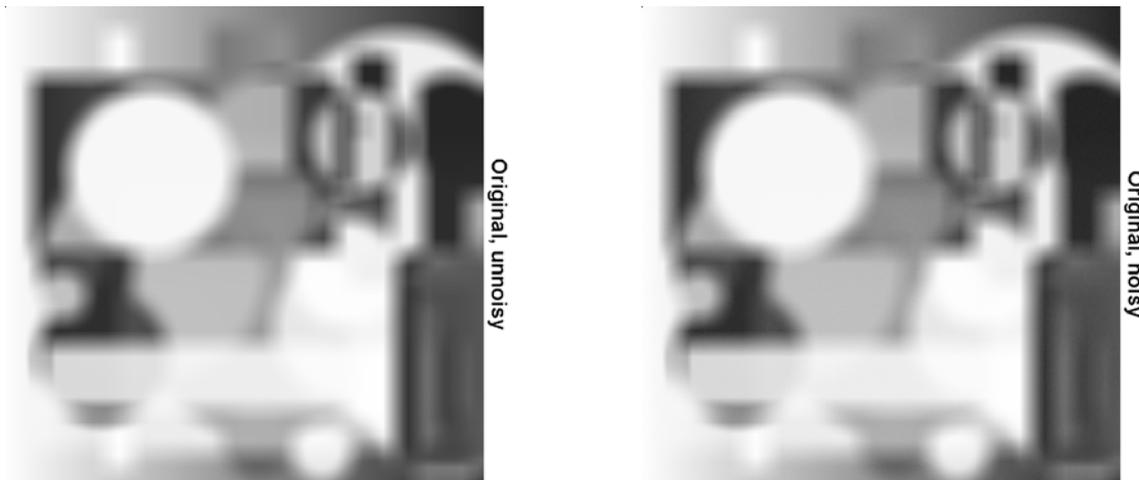


Fig. 2. Immagine simulata, prima (a sinistra) e dopo (a destra) la corruzione dell'immagine con rumore determinato dal processo di conteggio dei fotoni.

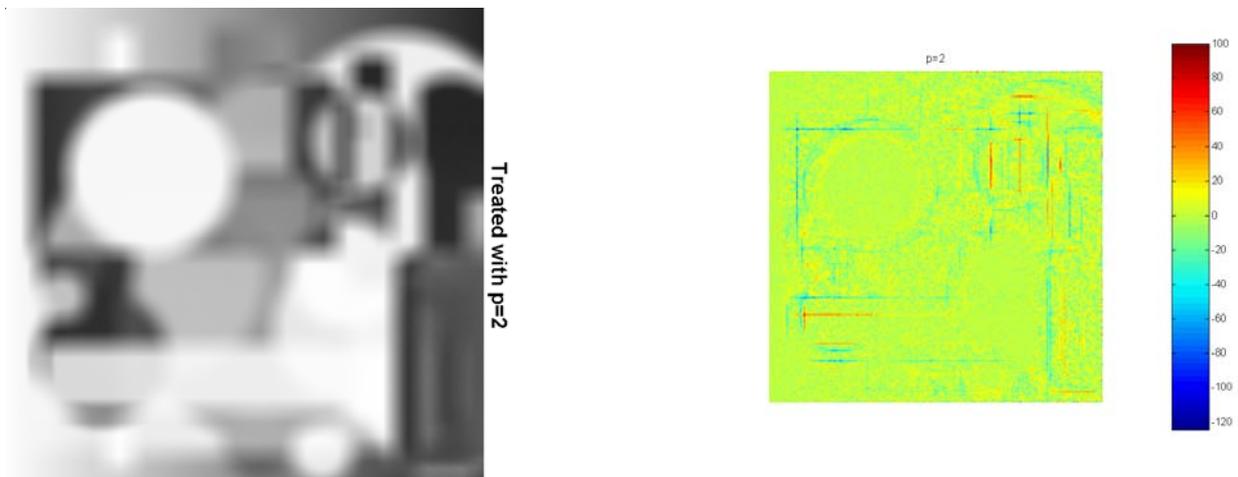


Fig. 3. Immagine trattata con $p=2$ (a sinistra) e differenza dell'immagine trattata con l'immagine originale, non rumorosa, mostrata in Fig. 2 a sinistra.

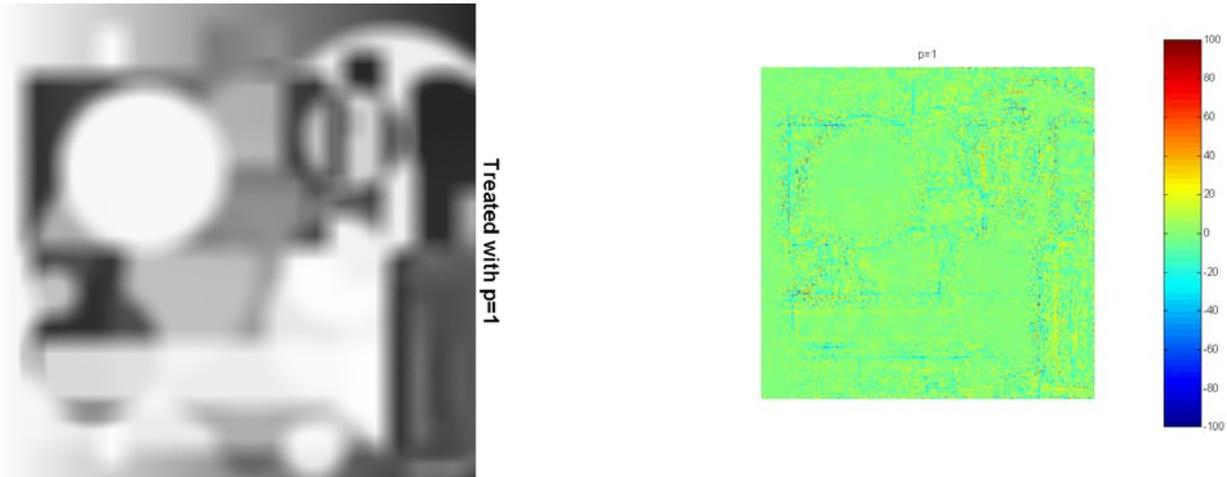


Fig. 4. Immagine trattata con $p=1$ (a sinistra) e differenza dell'immagine trattata con l'immagine originale, non rumorosa, mostrata in Fig. 2 a sinistra.

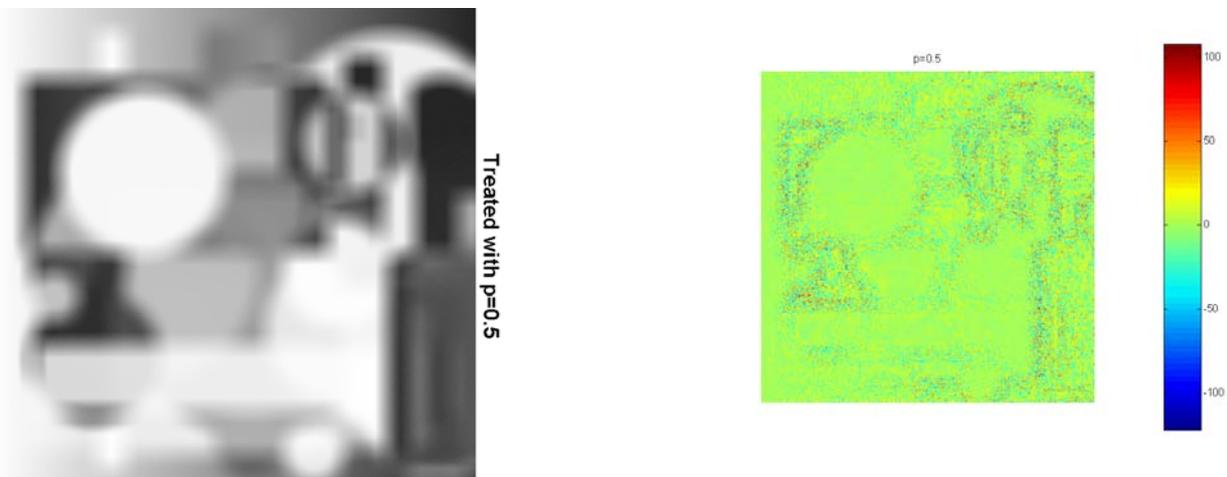


Fig. 5. Immagine trattata con $p=0.5$ (a sinistra) e differenza dell'immagine trattata con l'immagine originale, non rumorosa, mostrata in Fig. 2 a sinistra.

Risulta essere particolarmente significativa l'analisi dei pannelli di destra delle figure 3, 4 e 5, nei quali viene mostrata la differenza tra l'immagine trattata con il filtro di denoising e l'immagine originale priva di rumore.

Per $p=2$ (regolarizzazione alla Tikonov) si possono notare delle differenze significative tra l'immagine filtrata e l'immagine non rumorosa, concentrate soprattutto negli edge. Nelle zone uniformi dell'immagine vi è invece una differenza praticamente nulla. Ciò dimostra che l'approccio alla Tikonov è in grado di rimuovere in modo efficace il rumore; tuttavia, vi è un forte effetto di smoothing che tende a "rimuovere" gli edge dall'immagine.

Per $p=1$, l'effetto di rimozione degli edge è fortemente limitato, mentre permane la capacità del filtro di rimuovere il rumore. Lo stesso può essere detto per $p=0.5$, dove l'effetto di edge preserving è forse ancora più marcato.

Tali considerazioni permettono di scartare un approccio tipo Tikonov, non essendo dotato della proprietà edge preservino. Sono invece soluzioni valide quelle con $p \leq 1$. Dal momento poi che il caso più frequentemente trattato in letteratura è $p=1$ (TV), ci soffermeremo d'ora in poi solo su tale approccio.

Un ulteriore spunto di riflessione viene dall'analisi dell'approccio utilizzato per la minimizzazione della cifra di merito J , ovvero del metodo del gradiente. Tale metodo, implementato per una prima analisi dell'approccio al denoising mediante la teoria della regolarizzazione, è in realtà inadeguato per un'applicazione reale, in quanto la convergenza del metodo è molto lenta. Essendo necessario un grande numero di iterazioni (circa 100 negli esempi riportati sopra) per raggiungere la convergenza, non è possibile sviluppare filtri capaci di funzionare in tempo reale. E' dunque

necessario adottare un algoritmo di minimizzazione più efficiente, quali Expectation-Maximization (EM) o Scaled-Gradient (SG).

Ultimo spunto di riflessione, legato in parte all'efficienza computazionale, in parte alle prestazioni del filtro, riguarda il modo in cui viene calcolato il gradiente. Nelle equazioni utilizzate fino a questo punto, abbiamo considerato la norma L2 del gradiente, dove le componenti del gradiente sono state calcolate utilizzando due (W e N) degli otto vicini di ogni pixel. Dal punto di vista computazionale, potrebbe essere più efficiente utilizzare la norma L1 invece della norma L2; dal punto di vista delle prestazioni del filtro, potrebbe essere più performante calcolare il gradiente in modo più accurato, considerando un numero maggiore di vicini. Questi aspetti verranno trattati nelle prossime sezioni.

TV – come calcolare il gradiente

Una volta deciso di adottare TV come termine di regolarizzazione, è necessario specificare il modo in cui il gradiente viene calcolato sull'immagine. Vi sono in particolare due aspetti da prendere in considerazione:

- (i) quali degli otto vicini di ciascun pixel prendere in considerazione per il calcolo del gradiente; in teoria, i risultati migliori si dovrebbero ottenere con gradienti "centrati" nel pixel di interesse che prendono in considerazione il maggior numero possibile di vicini; d'altra parte il costo computazionale cresce al crescere del numero di vicini considerati.
- (ii) quale norma (L1 o L2) utilizzare per il calcolo del gradiente; in teoria, la norma L2 dovrebbe fornire risultati migliori di L1 in quanto la L2 è "isotropa" al contrario della L1; tuttavia, la norma L1 è computazionalmente più efficiente in quanto richiede solo il calcolo di due valori assoluti, mentre la norma L2 richiede il calcolo di due quadrati e di una radice quadrata.

Un ulteriore problema che nasce quando si considerano diversi modi per calcolare il gradiente è che il parametro di regolarizzazione μ deve essere opportunamente normalizzato per fare in modo che il termine di regolarizzazione J_R abbia lo stesso peso in J per la stessa immagine, quando J_R viene calcolato in modo differente. A tal fine, consideriamo per generalità un'immagine il cui gradiente possa assumere una qualsiasi direzione con la stessa probabilità; supponiamo poi che anche l'intensità del gradiente sia costante ed uguale a 1 su tutta l'immagine (tali ipotesi non sono restrittive). Se il gradiente viene calcolato utilizzando la norma L2, il valore medio della norma del gradiente sarà dato da:

$$\frac{1}{\pi/2} \cdot \int_0^{\pi/2} \|\nabla g(\omega)\|_2 d\omega = \frac{1}{\pi/2} \cdot \int_0^{\pi/2} \sqrt{(\sin \omega)^2 + (\cos \omega)^2} d\omega = \frac{1}{\pi/2} \cdot [\omega]_0^{\pi/2} = 1 \quad (22).$$

Nel caso invece in cui il gradiente sia calcolato in norma L1 avremo:

$$\frac{1}{\pi/2} \cdot \int_0^{\pi/2} \|\nabla g(\omega)\|_1 d\omega = \frac{1}{\pi/2} \cdot \int_0^{\pi/2} (|\sin \omega| + |\cos \omega|) d\omega = \frac{1}{\pi/2} \cdot [-\cos \omega + \sin \omega]_0^{\pi/2} = \frac{4}{\pi} \quad (23).$$

Dalla (22) e dalla (23) deriva che, se per un'immagine calcoliamo il gradiente utilizzando i due vicini N e W e la norma L2, con un parametro di regolarizzazione μ , e vogliamo trattare la stessa immagine calcolando il gradiente con gli stessi due vicini N e W, ma la norma L1, affinché il termine di regolarizzazione abbia lo stesso peso in tutti e due i casi dovremmo utilizzare in quest'ultimo caso un parametro di regolarizzazione pari a $(\pi/4) \cdot \mu$. Nel caso in cui utilizzassimo i quattro vicini N, E, S e W per il calcolo del gradiente in norma L1, dovremmo invece utilizzare un parametro di regolarizzazione pari a $(\pi/8) \cdot \mu$... Le costanti di normalizzazione per il μ possono essere ricavate in modo analogo per tutti gli altri casi di interesse.

Si tenga presente che, ogni volta che sia necessario calcolare un gradiente utilizzando due pixel a distanza diversa da 1, è necessario normalizzare il gradiente. Ad esempio:

$$g_{xy}(x, y) = \frac{g(x, y) - g(x-1, y-1)}{\sqrt{2}} \quad (24);$$

un altro caso interessante è quello in cui si considera il gradiente "centrato", ovvero:

$$g_x(x, y) = \frac{g(x+1, y) - g(x-1, y)}{2} \quad (25).$$

In tabella 1, vengono riportati i valori di μ da utilizzare per i casi più interessanti, al variare dei vicini utilizzati per il calcolo del gradiente e della norma scelta.

N	NE	E	SE	S	SW	W	NW	Norma	μ	Note
1	0	0	0	0	0	1	0	L2	$(1)\mu$	Caso "base"
1	0	1	0	1	0	1	0	L2	$(1)\mu$	Gradiente "centrato"
1	0	0	0	0	0	1	0	L1	$(\pi/4)\mu$	
1	0	1	0	1	0	1	0	L1	$(\pi/8)\mu$	
1	1	0	0	0	0	1	1	L1	$(\pi/8)\mu$	

Tabella 1. Normalizzazioni per il parametro di regolarizzazione m , al variare della norma e dei vicini impiegati per il calcolo del gradiente.

Abbiamo illustrato come normalizzare il parametro di regolarizzazione. Andremo nel prossimo paragrafo a vedere come effettuare la minimizzazione di J al variare della norma del gradiente e dei vicini utilizzati, utilizzando un algoritmo di minimizzazione più efficiente di quello del gradiente.

TV – come minimizzare con EM

L'algoritmo EM è noto per essere uno degli algoritmi più efficienti per la risoluzione di problemi tipo massima verosimiglianza / MAP. Di seguito ne forniamo una possibile derivazione (in prima istanza il metodo era stato derivato a partire dalla formula EM relativa alla ricostruzione in tomografia ed emissione, nel caso particolare in cui la matrice di proiezione del sistema coincida con l'identità).

Per la derivazione di EM, riscriviamo ancora una volta la cifra di merito che intendiamo minimizzare:

$$\left\{ \begin{array}{l} J_{DF}(g_{n,i} | g_i) = -\sum_{i=1}^N \{g_{n,i} \ln(g_i) - g_i\} \\ J_R(g_i) = \sum_{i=1}^N \|\nabla g_i\|_2^p \\ J(g_{n,i} | g_i) = J_{DF}(g_{n,i} | g_i) + \mu \cdot J_R(g_i) \\ J(\mathbf{g}_n | \mathbf{g}) = \sum_{i=1}^N J(g_{n,i} | g_i) \end{array} \right. \quad (26).$$

Per il pixel k -esimo, supponiamo che i vicini siano "costanti"; per ogni pixel, possiamo allora scrivere la funzione da massimizzare (passo di Expectation) come la J locale; ovvero, per ogni pixel k dell'immagine, vogliamo massimizzare $J(g_{n,k}|g_k)$, sotto l'ipotesi che i vicini siano costanti. Per effettuare tale massimizzazione (passo di Maximization) rispetto al valore di g_k , è sufficiente scrivere la derivata di $J(g_{n,k}|g_k)$ rispetto a g_k e porla uguale a zero, ovvero:

$$\frac{\partial J(g_{n,k} | g_k)}{\partial g_k} = \frac{\partial}{\partial g_k} \{-[g_{n,k} \ln(g_k) - g_k]\} + \mu \cdot \frac{\partial}{\partial g_k} J_R(g_k) = -\frac{g_{n,k}}{g_k} + 1 + \mu \cdot \frac{\partial}{\partial g_k} J_R(g_k) = 0 \quad (27).$$

La (26) non può essere risolta esplicitamente in modo generale, in quanto il termine g_k compare all'interno del termine di regolarizzazione (nel caso di norma L2, quindi sotto radice). E' però

possibile scrivere l'iterazione di punto fisso che permette di ricavare g_k , la quale costituisce anche l'equazione di aggiornamento per EM, cioè:

$$-\frac{g_{n,k}}{g_k} + 1 + \mu \cdot \frac{\partial}{\partial g_k} J_R(g_k) = 0 \Rightarrow \frac{g_{n,k}}{g_k} = 1 + \mu \cdot \frac{\partial}{\partial g_k} J_R(g_k) \Rightarrow g_k = \frac{g_{n,k}}{1 + \mu \cdot \frac{\partial}{\partial g_k} J_R(g_k)} \quad (28),$$

dove i termini a sinistra della (28) sono calcolati all'iterazione corrente, il termine a sinistra costituisce il valore aggiornato del k-esimo pixel. L'unico termine variabile a seconda della norma e dei vicini scelti per il calcolo della norma del gradiente è la derivata del termine di regolarizzazione. E' quindi sufficiente scrivere tale termine per i diversi casi da analizzare per avere automaticamente le equazioni di aggiornamento. Ricordiamo che, ove necessario per evitare divisioni per zero, deve essere utilizzata l'approssimazione di Hubert. Vediamo dunque, caso per caso, la forma di questo termine.

(i) Caso "base" – Gradiente non centrato calcolato con i pixel N ed W – Norma L2.

Gradiente:

$$\begin{aligned} \|\nabla g(x_i, y_i)\|_2 &= \sqrt{g_x(x_i, y_i)^2 + g_y(x_i, y_i)^2 + T^2} = \\ &= \sqrt{[g(x_i, y_i) - g(x_i - 1, y_i)]^2 + [g(x_i, y_i) - g(x_i, y_i - 1)]^2 + T^2} \end{aligned}$$

Parametro di regolarizzazione (normalizzato):

$$1 \cdot \mu$$

Derivata del termine di regolarizzazione:

$$\begin{aligned} \frac{\partial J_R(\mathbf{g})}{\partial g_k} &= \frac{\partial \sum_{i=1}^N \|\nabla g(x_i, y_i)\|_2}{\partial g_k} = \frac{\partial [\|\nabla g(x_k, y_k)\|_2 + \|\nabla g(x_k + 1, y_k)\|_2 + \|\nabla g(x_k, y_k + 1)\|_2]}{\partial g_k} = \\ &= \frac{[g_x(x_k, y_k) + g_y(x_k, y_k)]}{\|\nabla g(x_k, y_k)\|_2} - \frac{[g_x(x_k, y_k)]}{\|\nabla g(x_k + 1, y_k)\|_2} - \frac{[g_y(x_k, y_k + 1)]}{\|\nabla g(x_k, y_k + 1)\|_2} \end{aligned}$$

(ii) Gradiente "centrato" – Gradiente centrato calcolato con i pixel N, E, S ed W – Norma L2.

Gradiente:

$$\begin{aligned} \|\nabla g(x_i, y_i)\|_2 &= \sqrt{g_x(x_i, y_i)^2 + g_y(x_i, y_i)^2 + T^2} = \\ &= \sqrt{\left[\frac{g(x_i + 1, y_i) - g(x_i - 1, y_i)}{2} \right]^2 + \left[\frac{g(x_i, y_i + 1) - g(x_i, y_i - 1)}{2} \right]^2 + T^2} \end{aligned}$$

Parametro di regolarizzazione (normalizzato):

$$1 \cdot \mu$$

Derivata del termine di regolarizzazione:

$$\begin{aligned} \frac{\partial J_R(\mathbf{g})}{\partial g_k} &= \frac{\partial \sum_{i=1}^N \|\nabla g(x_i, y_i)\|_2}{\partial g_k} = \frac{\partial [\|\nabla g(x_k - 1, y_k)\|_2 + \|\nabla g(x_k, y_k + 1)\|_2 + \|\nabla g(x_k + 1, y_k)\|_2 + \|\nabla g(x_k, y_k - 1)\|_2]}{\partial g_k} = \\ &= \frac{[g_x(x_k - 1, y_k)]}{\|\nabla g(x_k - 1, y_k)\|_2} - \frac{[g_y(x_k, y_k + 1)]}{\|\nabla g(x_k, y_k + 1)\|_2} - \frac{[g_x(x_k + 1, y_k)]}{\|\nabla g(x_k + 1, y_k)\|_2} + \frac{[g_y(x_k, y_k - 1)]}{\|\nabla g(x_k, y_k - 1)\|_2} \end{aligned}$$

(iii) Gradiente non centrato calcolato con i pixel N ed W – Norma L1.

Gradiente:

$$\|\nabla g(x_i, y_i)\|_1 = |g_x(x_i, y_i)| + |g_y(x_i, y_i)| = |g(x_i, y_i) - g(x_i - 1, y_i)| + |g(x_i, y_i) - g(x_i, y_i - 1)|$$

Parametro di regolarizzazione (normalizzato):

$$\left(\frac{\pi}{4}\right) \cdot \mu$$

Derivata del termine di regolarizzazione:

$$\begin{aligned} \frac{\partial J_R(\mathbf{g})}{\partial g_k} &= \frac{\partial \sum_{i=1}^N \|\nabla g(x_i, y_i)\|_1}{\partial g_k} = \frac{\partial [\|\nabla g(x_k, y_k)\|_1 + \|\nabla g(x_k + 1, y_k)\|_1 + \|\nabla g(x_k, y_k + 1)\|_1]}{\partial g_k} = \\ &= \frac{\partial}{\partial g_k} [|g(x_k, y_k) - g(x_k - 1, y_k)| + |g(x_k, y_k) - g(x_k, y_k - 1)|] + \\ &\frac{\partial}{\partial g_k} [|g(x_k + 1, y_k) - g(x_k, y_k)| + |g(x_k + 1, y_k) - g(x_k + 1, y_k)|] + \\ &\frac{\partial}{\partial g_k} [|g(x_k, y_k + 1) - g(x_k - 1, y_k + 1)| + |g(x_k, y_k + 1) - g(x_k, y_k)|] = \\ &= \text{sign}[g_x(x_k, y_k)] + \text{sign}[g_y(x_k, y_k)] - \text{sign}[g_x(x_k + 1, y_k)] - \text{sign}[g_y(x_k, y_k + 1)] \end{aligned}$$

Sebbene i casi (i)-(iii) non esauriscano tutta la casistica possibile, sono indicativi per studiare la minimizzazione del funzionale J con EM.

E' da notare che l'applicazione della formula di aggiornamento (28) non porta ad una procedura convergente. Tale problema di EM era forse già stato notato da De Pierro (aspetto da approfondire). Per ovviare al problema è sufficiente inserire un parametro di damping, il cui valore viene determinato con un passo di line search, per forzare la convergenza dell'algoritmo. La pipeline risultante è illustrata in figura 6.

```

alpha=alpha_0; // Parametro di damping
for (niter=0; niter<NlTERS; niter++) // Iterazioni
{
    Jold=J(g); // Valore attuale della cifra di merito
    Deltag=gn/(1+mu*dJ/dg); // Direzione dello spostamento di g (EM)
    while [J(g+alpha*Deltag)>Jold] // Verifica della convergenza del passo (line search)
        alpha=alpha/2; // Aggiornamento del damping (line search)
    if (convergence) // Verifica della convergenza
        exit;
}

```

Fig. 6. Pipeline dell'algoritmo per la minimizzazione con damped EM.

Risultati sperimentali

Per confrontar tra di loro i diversi algoritmi di minimizzazione (metodo del gradiente e EM) abbiamo considerato la radiografia simulata mostrata in Fig. 2 e 3; dopo l'applicazione del filtro di denoising, abbiamo misurato le performance dei diversi metodi in termini di efficienza computazionale (numero di iterazioni richieste per la convergenza, tempo totale di calcolo per la convergenza, tempo per iterazione) e di qualità dell'immagine ricostruita (Mean Absolute Error, MAE; Root Mean Squared Error, RMSE; Kulbach Leibler Divergence, KLD). Per l'algoritmo EM la convergenza viene raggiunta quando il massimo cambiamento di livello di grigio all'interno dell'immagine è inferiore a 1 livello di grigio; per il metodo del gradiente, più lento, la convergenza è raggiunta per un cambiamento massimo di livello di grigio inferiore a 0.1. Il numero massimo di iterazioni per ogni algoritmo è stato fissato a 200. Inoltre, sono stati testati diversi parametri di regolarizzazione, per valori che vanno da 0.01 a 0.1.

Gli algoritmi testati sono:

Gradient: metodo del gradiente, norma L2 del gradiente non centrato calcolato utilizzando i vicini N ed W;

EM1: EM, norma L2 del gradiente non centrato calcolato utilizzando i vicini N ed W;

EM3: EM, norma L1 del gradiente non centrato calcolato utilizzando i vicini N ed W;

EM2: EM, norma L2 del gradiente centrato calcolato utilizzando i vicini N, E, S ed W;

EM5: EM, norma L1 del gradiente non centrato calcolato utilizzando i vicini N, E, S ed W;

EM7: EM, norma L1 del gradiente non centrato calcolato utilizzando i vicini W, NW, N, NE.

I risultati sono riportati in forma numerica in tabella 2 e in forma grafica nelle Fig. 7, 8, 9 e 10. Tutti gli algoritmi sono stati implementati e testati in Matlab.

GRADIENT							
Mu	eta	niter	time	MAE	RMSE	KLD	Sec per Iter
0.001	10	1	0.325576	15.5273	21.6205	45956.3	0.3256
0.01	10	200	44.1746	11.4847	16.2264	27451.2	0.2209
0.02	10	133	28.4358	11.3153	16.0511	26494.5	0.2138
0.05	10	58	12.3826	13.21	17.9761	36808.7	0.2135
0.1	10	41	9.19788	15.9538	20.7167	58627.8	0.2243

EM1							
Mu	damp	niter	time	MAE	RMSE	KLD	Sec per Iter
0.001	0.1	3	0.654518	15.1543	21.068	44852.6	0.2182
0.01	0.1	32	6.0824	8.95332	12.7147	21172	0.1901
0.02	0.1	40	7.55689	7.84755	11.4461	16490.4	0.1889
0.05	0.1	119	21.2975	7.52182	11.3491	13399.6	0.1790
0.1	0.1	200	35.0397	10.2095	17.5013	23080.8	0.1752

EM3							
Mu	damp	niter	time	MAE	RMSE	KLD	Sec per Iter
0.001	0.1	7	1.38207	14.7767	20.5206	43243.6	0.1974
0.01	0.1	43	8.04867	8.78368	12.4871	20652	0.1872
0.02	0.1	57	10.6492	7.87753	11.6617	16772.5	0.1868
0.05	0.1	200	35.4246	8.7166	13.1469	16525	0.1771
0.1	0.1	200	35.5276	12.8648	19.7018	29023.9	0.1776

EM2							
Mu	damp	niter	time	MAE	RMSE	KLD	Sec per Iter
0.001	0.1	1	0.336147	15.482	21.5464	46274.6	0.3361
0.01	0.1	37	6.76838	11.7893	16.395	31444	0.1829
0.02	0.1	35	6.61278	10.5679	15.0118	26367.5	0.1889
0.05	0.1	130	23.1872	10.0555	15.0721	22933.8	0.1784
0.1	0.1	200	35.3027	11.7744	19.4519	29233.6	0.1765

EM7							
Mu	damp	niter	time	MAE	RMSE	KLD	Sec per Iter
0.001	0.1	1	0.53867	15.419	21.4562	46001.6	0.5387
0.01	0.1	41	12.7559	9.31149	13.2299	22610.7	0.3111
0.02	0.1	66	20.0669	8.19079	12.0138	17571.5	0.3040
0.05	0.1	121	35.8209	8.28428	12.9752	15846.1	0.2960
0.1	0.1	200	57.9419	11.5289	19.1148	27442	0.2897

EM5							
Mu	damp	niter	time	MAE	RMSE	KLD	Sec per Iter
0.001	0.1	7	2.19573	14.7767	20.5206	43243.6	0.3137
0.01	0.1	43	12.6897	8.78368	12.4871	20652	0.2951
0.02	0.1	57	16.3814	7.87753	11.6617	16772.5	0.2874
0.05	0.1	200	54.6766	8.7166	13.1469	16525	0.2734
0.1	0.1	200	54.687	12.8648	19.7018	29023.9	0.2734

Tabella 2. Performance degli algoritmi sull'immagine simulata mostrata in Fig. 1 e 2.

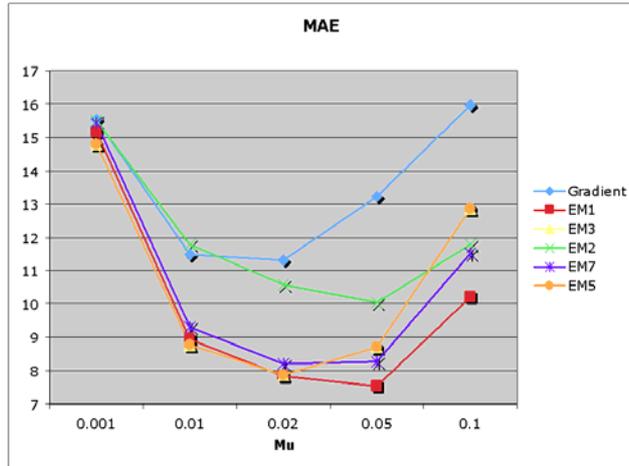


Fig. 7. MAE.

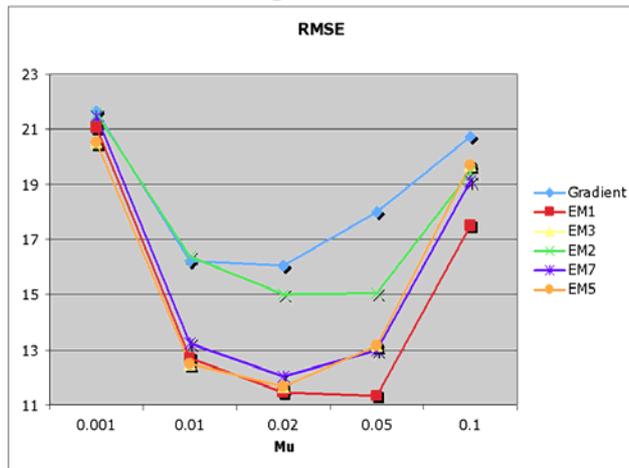


Fig. 8. RMSE.

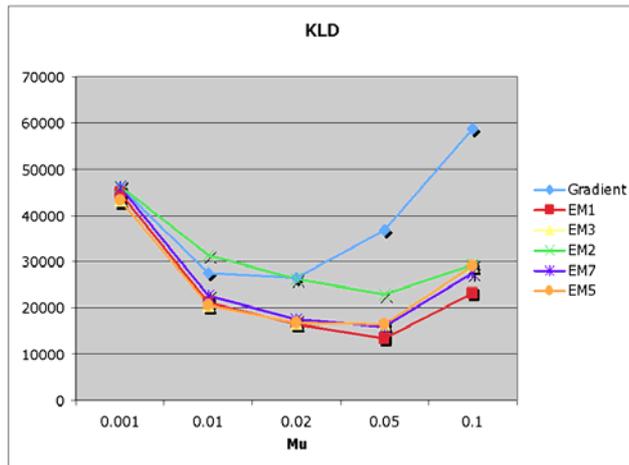


Fig. 9. KLD.

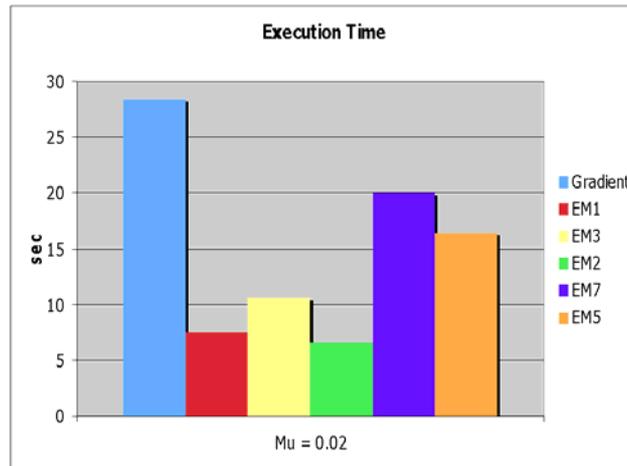


Fig. 10. Tempo di esecuzione per $\mu=0.02$.

Dall'analisi dei risultati risulta evidente che, quale che sia la definizione di J_R e l'algoritmo scelto per la minimizzazione, esiste un valore ottimale del parametro di regolarizzazione μ tale per cui l'effetto di denoising edge preserving dell'algoritmo è massimo (MAE, RMSE e KLD minimi). Per l'immagine considerata, tale valore si colloca tra 0.02 e 0.05, ed è stato determinato in maniera empirica. La scelta del parametro di regolarizzazione ottimale per ogni immagine resta un problema aperto, come discusso nella prossima sezione.

Il metodo del gradiente è, come atteso, il metodo più lento tra quelli implementati, dunque ha scarsa applicabilità pratica. Dal momento che ad ogni iterazione gli aggiornamenti sono piccoli, il metodo termina precocemente in alcuni casi, il che spiega tra l'altro le cattive performance dal punto di vista della qualità dell'immagine.

Un secondo metodo problematico dal punto di vista della qualità dell'immagine trattata è EM2. In questo caso il problema deriva dal modo in cui è stato definito il gradiente (gradiente centrato utilizzando i vicini N, E, S ed W). Tale gradiente risulta essere nullo per un'immagine tipo scacchiera; si osserva infatti, nelle immagini trattate con EM2, un artefatto a forma di scacchiera nelle zone omogenee, che peggiora notevolmente la qualità dell'immagine ricostruita. Quindi, contro quella che può sembrare una ragionevole intuizione iniziale, il gradiente centrato non porta in questo caso vantaggi. Le cose potrebbero cambiare nel momento in cui la norma del gradiente venisse calcolata utilizzando tutti gli otto vicini, in quanto la presenza dei gradienti diagonali lega in J_R i pixel vicini senza soluzione di continuità; tale argomento può essere oggetto di ricerche future.

Tutti gli altri metodi raggiungono performance simili dal punto di vista della qualità dell'immagine ricostruita; in particolare, EM5 risulta essere analogo a EM3, dal momento che la J_R è esattamente la stessa a parte un fattore 0.5 che viene compensato dalla normalizzazione di μ . Tra questi metodi, EM1 ed EM3 forniscono i risultati migliori dal punto di vista dell'efficienza computazionale e sono quindi i metodi più interessanti da tenere in considerazione per un'eventuale implementazione del metodo.

Alcuni risultati tipici ottenuti con EM3, 50 iterazioni, sono mostrati nelle figure seguenti. L'effetto più evidente si ha per immagini di tipo endoorale (intrinsecamente più rumorose).



Fig. 11. Una porzione (512x512) di immagine endoorale prima (a sinistra) a dopo (a destra) l'applicazione del filtro di denoising (gradiente non centrato in norma L1, calcolato utilizzando i vicini N ed W - $\mu=0.03$). Una gamma pari a 0.5 è stata applicata a posteriori.

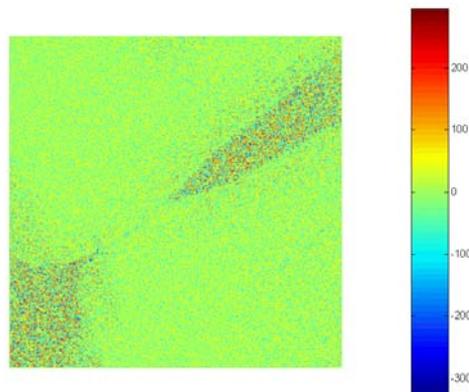


Fig. 12. Effetto del filtro di denoising per l'immagine mostrata in figura 11 (differenza tra immagine originale, rumorosa, e immagine trattata).

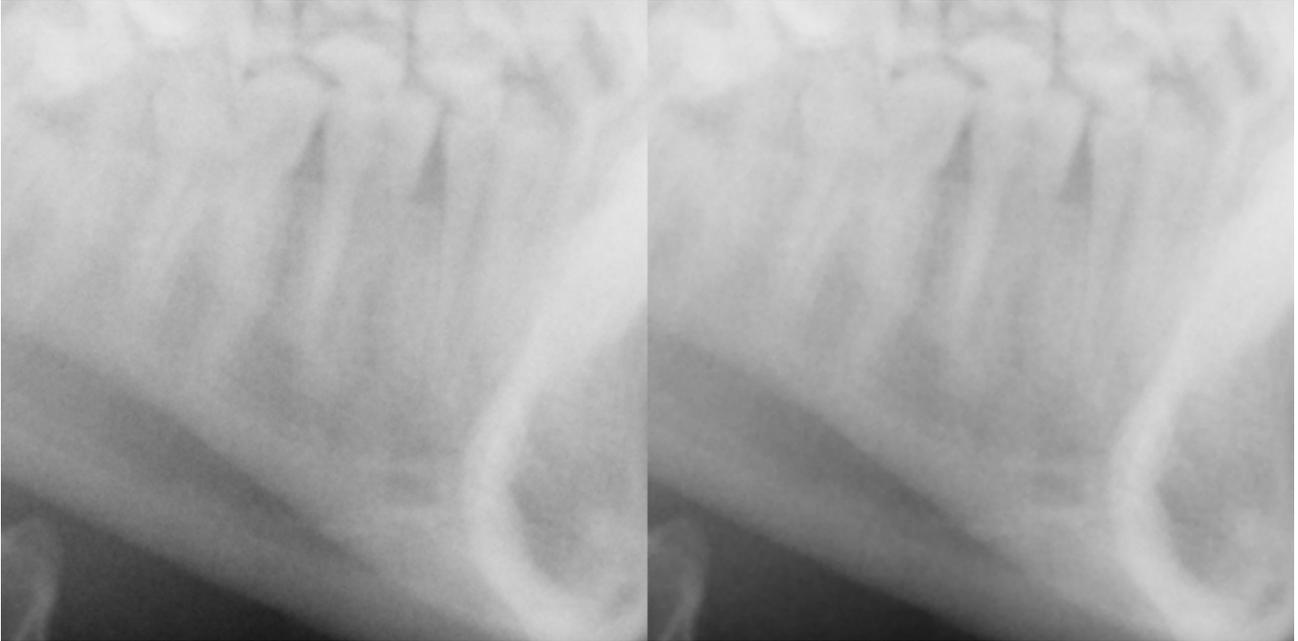


Fig. 13. Una porzione (512x512) di immagine cefalometrica prima (a sinistra) a dopo (a destra) l'applicazione del filtro di denoising (gradiente non centrato in norma L1, calcolato utilizzando i vicini N ed W – $\mu=0.01$). Una gamma pari a 0.5 è stata applicata a posteriori.

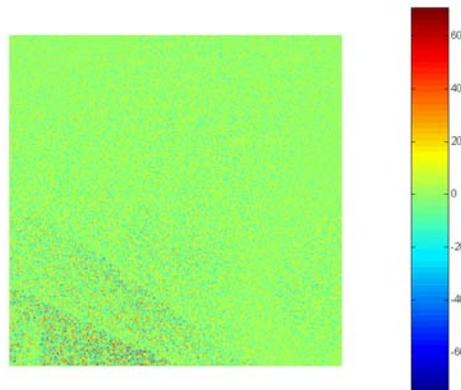


Fig. 14. Effetto del filtro di denoising per l'immagine mostrata in figura 13 (differenza tra immagine originale, rumorosa, e immagine trattata).

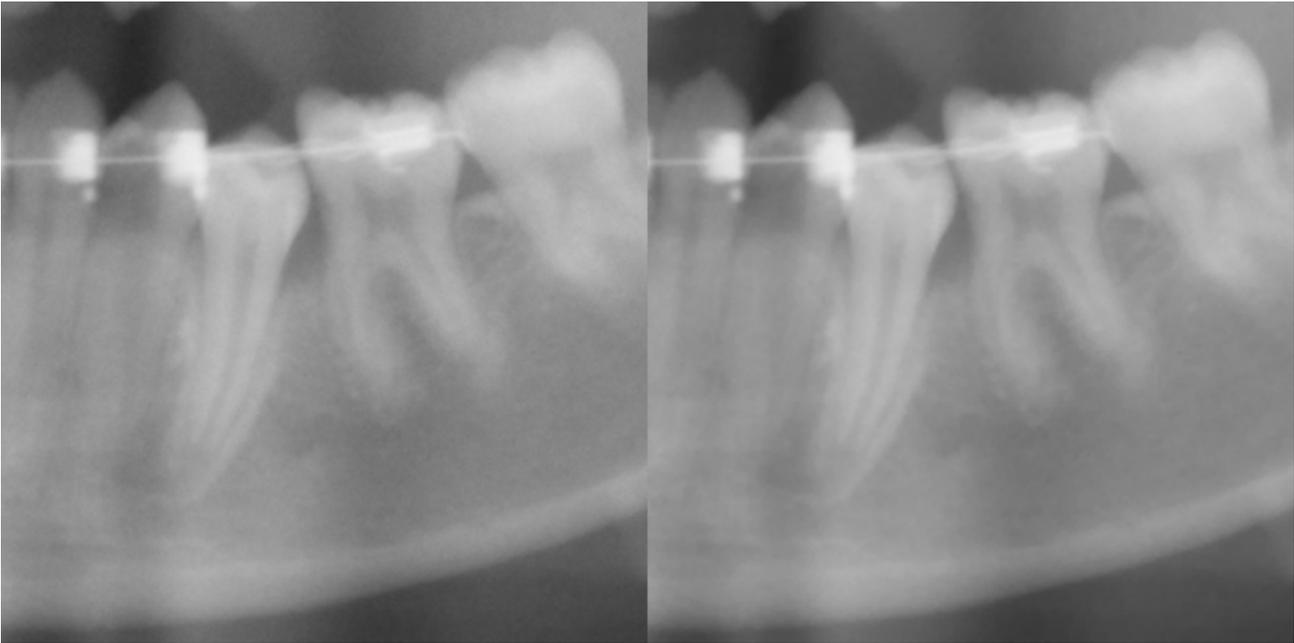


Fig. 13. Una porzione (512x512) di immagine panoramica prima (a sinistra) a dopo (a destra) l'applicazione del filtro di denoising (gradiente non centrato in norma L1, calcolato utilizzando i vicini N ed W - $\mu=0.01$). Una gamma pari a 0.5 è stata applicata a posteriori.

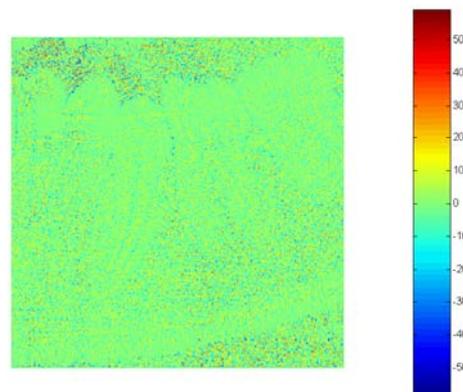


Fig. 16. Effetto del filtro di denoising per l'immagine mostrata in figura 11 (differenza tra immagine originale, rumorosa, e immagine trattata).

Conclusioni e problemi aperti

In conclusione, si è dimostrato che la teoria della regolarizzazione può essere applicata al problema del denoising in radiografia digitale per realizzare un filtro edge preserving funzionante in tempo reale.

Una delle ipotesi di base che è stata fatta riguarda la risposta all'impulso del sensore, che è stata supposta uguale ad un impulso. Tale ipotesi ha permesso di derivare in maniera semplice il termine di aderenza ai dati. Una modellizzazione più accurata della risposta all'impulso del sensore potrebbe portare un qualche miglioramento delle performance dell'algoritmo, il quale però risulterebbe più complicato dal punto di vista della formulazione matematica e più oneroso computazionalmente. Il problema diventerebbe un problema di denoising & deblurring. Visti i buoni risultati sperimentali raggiunti con l'approccio attuale, tale complicazione non sembra giustificata.

Per ottenere un algoritmo in grado di fornire una soluzione in tempi ragionevoli, è stato necessario lavorare sulla definizione del gradiente dell'immagine (norma L1 o norma L2, numero di vicini presi in considerazione) e sull'algoritmo di minimizzazione (Gradiente, EM, damped EM). Per quanto riguarda la formulazione del gradiente, sono rimasti alcuni casi da verificare i quali implicano però

l'utilizzo di un numero consistente di vicini, e dovrebbero essere quindi sfavorevoli dal punto di vista dell'impegno computazionale. Potrebbe comunque valere la pena di confrontare ogni possibilità per avere un quadro completo (in particolare, bisognerebbe considerare il gradiente centrato con 8 vicini, norma L2; e il gradiente non centrato con norma L2, 4 vicini inclusi i diagonali). Dal punto di vista dell'algoritmo di minimizzazione, abbiamo visto che il metodo del gradiente assicura la convergenza, ma in un tempo troppo lungo; EM non assicura la convergenza se non aggiungendo un passo di line search (damped EM). Si veda anche la Fig. 17. L'efficienza computazionale di damped EM è comunque tale da garantire un'implementazione in tempo reale (o quasi) dell'algoritmo. Un'ulteriore miglioramento potrebbe verificarsi andando a considerare metodi di minimizzazione tipo Scaled Gradient (SG); alcuni risultati sperimentali preliminari vanno in questo senso.

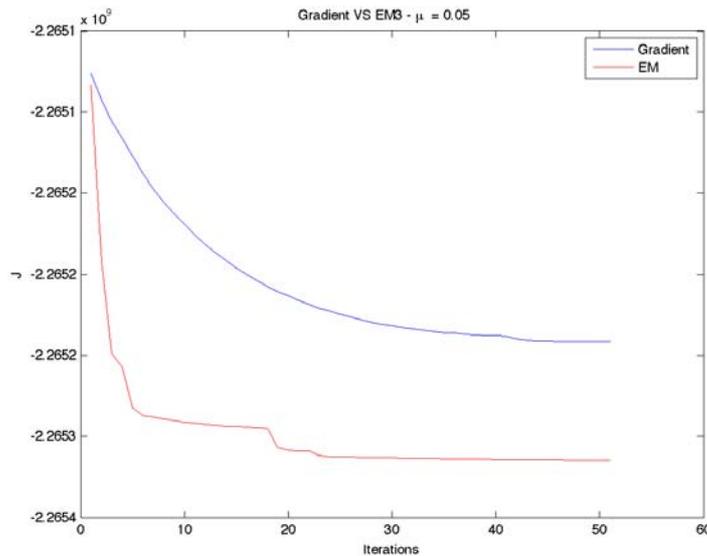


Fig. 17. Velocità di convergenza (discesa di J) per l'algoritmo EM (damped) in confronto al metodo del gradiente, per $m=0.02$ e l'immagine simulata di Fig. 1 e 2.

Un problema aperto riguarda la stima ottimale del parametro di regolarizzazione. Tale parametro regola il peso del termine J_R rispetto al termine J_{DF} nella J da minimizzare. Il parametro μ ottimale dipende dal contenuto in frequenza e dal livello di rumorosità di un'immagine, quindi dal gain del sensore (più il gain è alto, più l'immagine è rumorosa). E' evidente, quindi, che il parametro ottimale per un'immagine endoorale non può essere lo stesso per un'immagine panoramica o ceflaometrica. Tale parametro può essere determinato in maniera empirica per ogni tipo di immagine da trattare (ad esempio, per le endoorali, un valore attorno a 0.05 sembra ragionevole). Uno spunto di ricerca può riguardare il calcolo automatico del parametro di regolarizzazione attuale. Un ulteriore spunto può riguardare lo sviluppo di un parametro di regolarizzazione adattativo.

Allegati

In allegato alla presente relazione:

codice .m;
immagini .tif (risultati tipici).