

IST. EL. INF.  
BIBLIOTECA  
P. 001. **ARCHIVIO**

*Consiglio Nazionale delle Ricerche*

**ISTITUTO DI ELABORAZIONE  
DELLA INFORMAZIONE**

**PISA**

**MODELLI A CAMPI DI MARKOV PER LA RICOSTRUZIONE  
BAYESIANA DI IMMAGINI**

**LUIGI BEDINI E ANNA TONAZZINI**

Nota Interna B4 - 29  
Giugno 1989

**MODELLI A CAMPI DI MARKOV PER LA  
RICOSTRUZIONE BAYESIANA DI IMMAGINI**

*Luigi Bedini e Anna Tonazzini*

Istituto di Elaborazione della Informazione

Consiglio Nazionale delle Ricerche

Via S.Maria, 46  
56100 PISA

## 1. Introduzione

Negli ultimi anni si è venuta formando la convinzione che molti dei problemi già risolti o che costituiscono ancora oggetto di studio nei settori della computer vision e della ricostruzione di immagini presentano un aspetto unificante, costituito dalla loro possibile formulazione in termini di problemi inversi mal-posti [1,2,3,4]. In letteratura è stato infatti ampiamente discusso come problemi quali la stereopsi, l'edge detection, la determinazione del flusso ottico in computer vision [5,6,7,33], e la deconvoluzione, la superrisoluzione, la tomografia, per citare alcune applicazioni nel settore della ricostruzione di immagini [23,34], sono formulabili come problemi inversi e presentano spesso infinite soluzioni o sono, quantomeno, sempre mal-condizionati.

La definizione più generale di un problema inverso mal-posto è data con riferimento alla relazione funzionale

$$g=O[f] \tag{1.1}$$

dove  $f$  è un elemento di uno spazio  $X$ ,  $g$  è un elemento di uno spazio  $Y$  ed  $O$  è un operatore noto da  $X$  in  $Y$ . Su questa base si definisce "problema diretto" il problema di determinare  $g$  (soluzione) noto  $f$  (dati), mentre si definisce "problema inverso" il problema di ricavare  $f$  (soluzione) noto  $g$  (dati).

Un problema formulabile secondo l'eq. (1,1), sia esso diretto o inverso, è detto ben-posto quando, secondo la definizione proposta da Hadamard nel 1923, soddisfa ciascuna delle tre seguenti condizioni [22,27]:

- 1) la soluzione è unica
- 2) la soluzione esiste per ogni valore dei dati
- 3) la soluzione dipende in modo continuo dai dati

In generale, poi, affinché sia matematicamente trattabile, un problema oltre che ben-posto deve essere anche ben-condizionato, cioè la soluzione deve essere robusta (stabile) nei confronti del rumore eventualmente presente sui dati. In pratica quello che si richiede è che piccole perturbazioni sui dati si ripercuotano in piccole perturbazioni sulla soluzione. La condizione 3 è una condizione necessaria ma non sufficiente alla stabilità del problema [27].

Usualmente un problema diretto, sia esso puramente matematico o derivato da una legge fisica, è ben-posto, e quindi ammette una sola soluzione, univocamente determinabile sulla base della conoscenza dell'operatore  $O$ . I problemi dell'ottica classica forniscono tipici esempi di problemi diretti ben-posti.

Viceversa, la maggior parte dei problemi inversi, anche se corrispondenti a problemi diretti ben-posti, sono mal-posti, nel senso che non soddisfano una o più delle tre condizioni proposte da Hadamard. In pratica, può non essere soddisfatta la condizione 1, che richiede che l'operatore  $O$  sia invertibile, cioè che ammetta un operatore inverso  $O^{-1}$ , oppure la condizione 2, che richiede che  $O^{-1}$  sia definito ovunque in  $Y$ , o la condizione 3, che richiede che  $O^{-1}$  sia continuo.

L'idea fondamentale per la regolarizzazione di problemi inversi mal-posti, cioè per il recupero di una soluzione unica e stabile al rumore, è quella di restringere la classe delle soluzioni ammissibili imponendo vincoli derivanti dalle conoscenze a priori che si hanno sul problema e sulla soluzione desiderata, e quindi di scegliere un funzionale costo da ottimizzare su tale classe.

L'approccio più tradizionale alla regolarizzazione è quello deterministico in cui i vincoli sono rappresentati da proprietà note della soluzione (positività, estensione finita, limitatezza, ecc.), e il funzionale costo normalmente rappresenta una certa misura globale (entropia, energia, ecc.) che deve essere ottimizzata dalla soluzione desiderata [1,22,23,29,30,31].

Un approccio alternativo è quello probabilistico in cui i vincoli sono sostanzialmente rappresentati da una descrizione della soluzione in termini di una assegnata densità di probabilità "a priori", mentre il funzionale costo è determinato sulla base della regola di Bayes, che lega la probabilità a priori alla probabilità "a posteriori", cioè alla probabilità di avere una certa soluzione, noti i dati disponibili. In questo modo il problema inverso originale è ricondotto al calcolo della stima che massimizza tale probabilità a posteriori [11,12], o che minimizza il valore atteso (rispetto alla probabilità a posteriori) di una appropriata funzione errore [14,17,32].

Sia nell'approccio deterministico che nell'approccio bayesiano, comunque, il problema è formulato in termini di ottimizzazione vincolata, e può essere risolto con tecniche che sono indipendenti dal tipo di approccio adottato. In particolare, se tale problema risulta convesso, e questo è il caso più semplice, esistono svariate tecniche di risoluzione efficienti (differenziazione, algoritmi di descent)[19,20,21]. Se il problema è invece non convesso, presenta cioè dei minimi locali, le cose si complicano notevolmente, e le tecniche di risoluzione, ancora oggetto di indagine da parte degli addetti ai lavori, sono sostanzialmente classificabili in due categorie: tecniche deterministiche, che approssimano la soluzione come limite di sequenze di soluzioni di problemi convessi (algoritmo GNC) [8,9,10]; tecniche probabilistiche, o di rilassamento stocastico, che ispezionano il funzionale costo su sequenze di punti generati casualmente (metodo Monte Carlo, algoritmo di Metropolis, annealing simulato, campionatore di Gibbs) [12,15,18].

In questo lavoro verrà preso in esame in dettaglio l'approccio bayesiano al problema inverso della ricostruzione di immagini. Il problema verrà formulato con riferimento all'eq.(1,1), in cui  $f$  rappresenterà di volta in volta l'immagine da ricostruire o l'oggetto da stimare (soluzione del problema), mentre  $g$  rappresenterà le misure disponibili o l'immagine distorta (dati del

problema). In particolare verrà studiata la possibilità di utilizzare un modello probabilistico per la soluzione basata sui Markov Random Fields e sulle distribuzioni di Gibbs, e verrà affrontato il problema della massimizzazione della probabilità a posteriori mediante algoritmi di rilassamento stocastico.

## 2. L'approccio Bayesiano nella ricostruzione di immagini

Nell'approccio probabilistico alla ricostruzione l'immagine da stimare è modellata come un insieme di variabili casuali,  $F=\{F_{ij}\}$ , a valori  $f=\{f_{ij}\}$  in un insieme  $\Lambda$  (spazio dei valori o degli stati), e ciascuna associata idealmente ad un sito di uno spazio geometrico  $S=\{s_{ij}\}$  (spazio dei siti). L'insieme

$$\Omega = \{f : f_{ij} \in \Lambda\}$$

è detto spazio delle configurazioni.

Nella formulazione più semplice lo spazio dei siti  $S$  è costituito dai punti di un array bidimensionale discreto (pixel), e lo spazio dei valori  $\Lambda$  è costituito da un insieme discreto e finito di numeri (usualmente i numeri interi fra 0 e 255), che rappresentano i valori di intensità associabili ad ogni singolo pixel. Si parlerà allora di processo di intensità per l'immagine  $F$  e di siti di intensità per l'insieme  $S$ . Si vedrà però in seguito come il modello per l'immagine possa essere complicato per tenere conto di altri attributi geometrici dell'immagine stessa, tipicamente gli elementi di edge, o elementi di linea.

Il problema della ricostruzione di immagini si configura in generale come il problema di determinare la miglior stima  $f$  del processo  $F$ , data una sua versione distorta ed affetta da rumore  $g$ .

Il modello di degradazione assunto è usualmente espresso dalla seguente relazione

$$G = \Phi(H(F)) \oplus N$$

che particularizza l'eq. (1,1), e dove il termine  $H(F)$  rappresenta l'uscita di un sistema o di un operatore,  $\Phi$  rappresenta una trasformazione possibilmente non lineare, tipicamente associabile al processo di registrazione,  $\oplus$  rappresenta addizione o moltiplicazione,  $N$  è l'insieme di variabili casuali associabili al rumore e  $G$  è l'insieme di variabili casuali associabili ai dati. Per semplicità di trattazione considereremo nel seguito il caso particolare in cui  $\Phi$  è la trasformazione identica e  $\oplus$  è l'addizione, cioè

$$G = H(F) + N \quad (2.1)$$

I dati a disposizione,  $g$ , rappresenteranno una particolare realizzazione del processo  $G$ , e la (2.1) potrà essere riscritta come:

$$g = H(f) + n$$

Così il problema della ricostruzione di immagini è formulabile come il problema di stimare  $f$  dato  $g$ .

Per la formula di Bayes vale:

$$P(F=f | G=g) = \frac{P(G=g | F=f) P(F=f)}{P(G=g)} \quad (2.2)$$

da cui scaturisce l'idea di scegliere come soluzione al problema la realizzazione  $f$  di  $F$  che massimizza la probabilità a posteriori di avere  $f$  dato  $g$ , cioè:

$$\max_f \frac{P(G=g | F=f) P(F=f)}{P(G=g)} \quad (2.3)$$

Assumendo  $P(G=g)$  costante, in quanto l'evento  $g$  si è verificato (dati a disposizione), e nell'ipotesi di conoscere  $P(F=f)$ , essendo  $P(G=g|F=f)$  determinabile sulla base della relazione che lega fra loro  $F$  e  $G$ , il problema di stimare  $f$  è completamente ricondotto al problema di ottimo

$$\max_f P(G=g|F=f) P(F=f) \quad (2.4)$$

Questo tipo di stima prende il nome di stima MAP (Maximum A Posteriori estimate).

Talvolta non è possibile fare ipotesi sulla probabilità a priori per  $F$ , cioè non si è in grado di imporre vincoli sulla soluzione desiderata. In questo caso quello che si fa è di accettare che tutte le possibili configurazioni per  $F$  siano equiprobabili, cioè che  $P(F=f)$  sia uniforme. Allora la stima MAP si riduce alla cosiddetta stima ML (Maximum Likelihood estimate) formulabile come:

$$\max_f P(G=g|F=f) \quad (2.5)$$

Nell'ipotesi che il processo rumore  $N$  sia indipendente da  $F$  e abbia densità di probabilità gaussiana con media nulla (per semplicità) e matrice di covarianza costante pari a  $\sigma^2 I$ , si può ricavare una semplice espressione per  $P(G=g|F=f)$ . Si ha infatti:

$$\begin{aligned} P(G=g | F=f) &= P(H(F) + N=g | F=f) = \\ &= P(N=g-H(F)|F=f) = P(N=g-H(f)) = \\ &\exp \left[ -\frac{\|g-H(f)\|^2}{2\sigma^2} \right] \end{aligned} \quad (2.6)$$

a meno di costanti moltiplicative.

In questo modo la stima MAP si riduce a:

$$\max_f \exp \left[ -\frac{\|g-H(f)\|^2}{2\sigma^2} \right] P(F=f) \quad (2.7)$$

mentre la stima ML diventa:

$$\max_f \exp \left[ -\frac{\|g-H(f)\|^2}{2\sigma^2} \right] \quad (2.8)$$



equivalente a

$$\min_f \frac{\|g-H(f)\|^2}{2\sigma^2} \quad (2.9)$$

che coincide in pratica con la stima prodotta dal metodo del filtro inverso, definito quando  $H$  è invertibile.

E' opinione comune che la formulazione MAP e l'approccio Bayesiano in genere siano particolarmente adatti a risolvere il problema della ricostruzione di immagini, anche perchè non richiedono la conoscenza di  $P(G=g)$ , in generale molto complessa. Per contro tali formulazioni comportano un problema computazionale molto gravoso, dovuto sia al numero delle variabili casuali in gioco (almeno tante quanti sono i pixel nell'immagine), sia al fatto che la risultante funzione obiettivo può presentare più massimi locali.

Le tecniche deterministiche per la risoluzione di problemi MAP si basano sulla generazione di sequenze di immagini che aumentano monotonamente il valore della probabilità a posteriori. Tali metodi possono però bloccarsi in corrispondenza di massimi locali.

Altre tecniche, dette di rilassamento stocastico, consentono cambi nell'immagine stimata che possono anche decrescere il valore della probabilità a posteriori. Questi cambi sono fatti in modo casuale e hanno come scopo quello di evitare la convergenza a massimi locali. Rimane tuttavia il problema computazionale dovuto all'esigenza di estrarre un campione di un processo di dimensione in genere molto elevata.

### 3. La scelta della probabilità a priori

Come si è visto, nelle stime Bayesiane la soluzione dipende solo dalle conoscenze a priori che si hanno sulle probabilità di  $F$  e

di  $N$  e dai dati  $g$ . Il problema fondamentale è costituito quindi dalla scelta del modello probabilistico per la soluzione.

Una scelta comune è quella di utilizzare una probabilità a priori di tipo gaussiano, cioè del tipo

$$P(F=f) = \exp[-1/2 (f-f_0)^T R_f^{-1} (f-f_0)] \quad (3.1)$$

a meno di costanti moltiplicative.

Tale scelta comporta una notevole semplificazione del funzionale costo MAP, ed inoltre essa risulta fisicamente plausibile [11]. Infatti, verosimilmente, le immagini del mondo reale sono sufficientemente varie da far sì che un modello di tipo gaussiano ne possa descrivere un ben specificato sottoinsieme. Tale sottoinsieme sarà ovviamente caratterizzato dalla scelta di  $f_0$  (la media) e di  $R_f$  (la matrice di covarianza).

Ad esempio, mediante la scelta di una media non uniforme, ma di una covarianza stazionaria, si potrà modellare abbastanza bene la classe delle immagini di volti umani tutti ripresi nella stessa posizione. Così  $f_0$  (immagine non costante) rappresenterà le strutture di base di tali immagini (ovale del volto, posizione di occhi, naso, bocca, ecc.), sfrondate dai dettagli, cioè viste come features geometriche, mentre la matrice di covarianza  $R_f$  regolerà le fluttuazioni casuali attorno a tale media, per dare i dettagli di ogni singola immagine e quindi le differenze fra immagine ed immagine.

Si avrà quindi che la stima MAP corrispondente ad una scelta gaussiana per  $F$  sarà data da:

$$\max_f \exp \left[ -\frac{\|g-H(f)\|^2}{2\sigma^2} \right] \exp \left[ -\frac{1}{2} (f-f_0)^T R_f^{-1} (f-f_0) \right] \quad (3.2)$$

o equivalentemente

$$\min_f \frac{\|g-H(f)\|^2}{2\sigma^2} + \frac{1}{2} (f-f_0)^T R_f^{-1} (f-f_0) \quad (3.3)$$

Esistono particolari scelte del modello probabilistico per  $F$  per cui il problema MAP che ne deriva può essere messo in stretta relazione coi classici metodi di ottimizzazione vincolata deterministici. Tale relazione è in genere puramente formale, in quanto la filosofia che sta dietro ai due approcci è completamente diversa. Tuttavia questo parallelismo può risultare utile ai fini dell'individuazione della tecnica di risoluzione [16], in quanto mette in evidenza come algoritmi di risoluzione ideati per un approccio possano essere utilizzati anche nell'altro.

Un esempio di quanto detto sopra è stato dato nel paragrafo precedente in cui si è visto come, scegliendo una densità di probabilità a priori di tipo uniforme per  $F$ , il funzionale della stima ML risulta formalmente identico al funzionale costo del metodo del filtro inverso.

Un altro esempio è fornito dalla scelta di  $P(F=f)$  del seguente tipo:

$$P(F=f) = \exp[-\beta f^T \log f] \quad (3.4)$$

che evidenzia come il problema di ricostruzione di immagini mediante il metodo della Massima Entropia può essere formulato come un problema MAP [16]. Innanzi tutto occorre osservare che un tale tipo di distribuzione a priori sembra essere ragionevole per immagini di tipo impulsivo (ad esempio fotografie di campi di stelle) e ciò ha riscontro con il fatto che il metodo di ricostruzione di massima entropia è stato principalmente applicato a questo tipo di immagini [25]. La formulazione del problema MAP per una tale scelta diventa:

$$\max_f \exp[-\|g-H(f)\|^2/2\sigma^2] \exp[-\beta f^T \log f] \quad (3.5)$$

oppure

$$\min_f \beta f^T \log f + \|g-H(f)\|^2/2\sigma^2 \quad (3.6)$$

La tipica formulazione del metodo di massima entropia è invece la seguente:

$$\max_f -f^T \log f \quad \text{soggetto al vincolo} \quad \|g-H(f)\|^2 = \sigma^2$$

che, con l'uso della teoria dei moltiplicatori di Lagrange, risulta equivalente al problema di ottimizzazione non vincolata:

$$\min_f f^T \log f + \lambda (\|g-H(f)\|^2 - \sigma^2) \quad (3.7)$$

per un opportuno valore di  $\lambda$  [19,26,31].

E' evidente l'analogia fra i funzionali costo dei due metodi. Questi avranno inoltre lo stesso valore di minimo scegliendo, nel funzionale MAP,  $\beta=1/2\lambda\sigma^2$ . Occorre tuttavia ribadire che i due metodi sono concettualmente diversi, nel senso che è diversa l'interpretazione che si può dare delle soluzioni ottenute.

Nella caratterizzazione probabilistica per  $F$ , quello che si vorrebbe è trovare una classe di modelli che ben si adattino a rappresentare una categoria abbastanza ampia di immagini realistiche e legate fra loro da caratteristiche comuni. In altre parole il modello dovrebbe essere in grado di incorporare le conoscenze a priori che si possono avere sulla classe delle soluzioni ammissibili per il problema di ricostruzione, prime fra tutte le caratteristiche tipiche di un'immagine, quali, ad esempio, le discontinuità. Le proprietà di una tale classe ideale di modelli sono state così riassunte [14]:

1. La dipendenza probabilistica fra gli elementi del processo dovrebbe essere locale, in modo da modellare superfici (immagini) che sono uniformi solo a tratti. Questa proprietà fa sì che l'algoritmo di ricostruzione risulti distribuito e quindi implementabile efficacemente in modo parallelo.

2. La classe dovrebbe essere ricca a sufficienza da poter modellare un'ampia varietà di comportamenti qualitativamente differenti.
3. La relazione fra i parametri del modello e le caratteristiche dei corrispondenti campioni del processo dovrebbe essere trasparente, in modo da rendere facile la specificazione dei modelli stessi.
4. Dovrebbe essere possibile rappresentare esplicitamente la densità di probabilità a priori per l'immagine da ricostruire, in modo che la teoria di Bayes possa essere applicata.
5. Dovrebbe essere possibile individuare procedure efficienti, sia per generare campioni dalle densità di probabilità specificate, in modo da verificare che il modello realmente fitti le conoscenze a priori, sia per calcolare stimatori ottimi.

Una classe di modelli probabilistici che soddisfa le richieste specificate nei cinque punti sopra è quella dei Markov Random Fields (MRF) su reticoli finiti [24].

Un MRF ha la proprietà che la probabilità condizionale di un elemento rispetto a tutti gli altri dipende solo da un opportuno intorno dell'elemento stesso. Questa proprietà soddisfa in particolare la richiesta 1. Un'altra proprietà molto importante di un MRF, evidenziata in [12], è che la densità di probabilità congiunta delle configurazioni del campo può sempre essere espressa in termini di una distribuzione di Gibbs. La forma assunta da una distribuzione di Gibbs si presta molto bene a rappresentare e ad esplicitare la dipendenza locale fra gli elementi del campo (richieste 1, 3, 4). Inoltre si prova che, nell'ipotesi che il rumore sia gaussiano, la distribuzione a posteriori può a sua volta essere rappresentata come una distribuzione di Gibbs. Questo, insieme con la relativa semplicità di derivare le singole probabilità

condizionali, consente di individuare efficienti algoritmi di stima (richiesta 5).

#### 4. Markov Random Fields e Distribuzioni di Gibbs

La modellizzazione di un processo immagine come un MRF è ampiamente utilizzato in letteratura, in quanto è fisicamente plausibile per un'immagine l'assunzione che il valore di un dato pixel dipenda dai valori dei pixel ad esso adiacenti, secondo un assegnato sistema di intorni. Tale approccio, di per se stesso, presenta comunque alcune difficoltà, comuni a tutti gli approcci di tipo Bayesiano. Infatti, in generale, data la probabilità congiunta  $P(\mathbf{X}=\omega)$  di un processo, non è immediato ricavare le densità di probabilità condizionali ad esso associate. Viceversa, dato un insieme di funzioni del tipo  $P(X_s=x_s | X_r=x_r, r \neq s)$ , è estremamente difficile stabilire se esse possono essere viste come probabilità condizionali per una densità di probabilità congiunta (necessariamente unica) [12].

Tuttavia, come già accennato, l'esplicitazione dell'equivalenza fra MRF e distribuzioni di Gibbs fornisce un valido strumento per superare le sopradette difficoltà. Innanzi tutto risulta possibile derivare una formula esplicita per la densità di probabilità congiunta in termini di una *funzione energia* la cui scelta, insieme con la scelta del sistema di intorni, rappresenta uno strumento potente per modellare le continuità spaziali locali e altre caratteristiche della scena stessa. Inoltre, dato il modello di degradazione assunto (2.1), nell'ipotesi che il rumore sia gaussiano e con media e varianza note, anche la probabilità congiunta a posteriori è ancora esprimibile come una distribuzione di Gibbs, e risulta estremamente semplice ricavare da essa le probabilità condizionali associate.

Sia  $\mathbf{X}$  un insieme di  $N$  variabili aleatorie, in corrispondenza biunivoca con un insieme  $S$  di  $N$  elementi detto insieme dei siti, cioè

$$X = \{ X_s, s \in S \} \quad (4.1)$$

Si consideri per semplicità, anche se non necessario, che  $X$  sia isotropico e omogeneo e che ogni variabile aleatoria assuma valori discreti in un comune spazio degli stati

$$\Lambda = \{0, 1, 2, \dots, K-1\} \quad (4.2)$$

Sia  $\Omega$  l'insieme di tutte le possibili configurazioni

$$\Omega = \{ \omega = (x_{s_1}, x_{s_2}, \dots, x_{s_N}) : x_{s_i} \in \Lambda, i=1, 2, \dots, N \} \quad (4.3)$$

Su  $S$  si definisca un sistema di intorni

$$\mathcal{G} = \{ \mathcal{G}_s, s \in S \} \quad (4.4)$$

con le proprietà che:

1.  $s \in \mathcal{G}_s$
  2.  $s \in \mathcal{G}_s \Leftrightarrow r \in \mathcal{G}_s$
- (4.5)

Allora  $X$  è un Markov Random Field (MRF) rispetto alla coppia  $(S, \mathcal{G})$  se valgono le seguenti proprietà [24]:

$$P(X=\omega) > 0 \quad \forall \omega \in \Omega$$

$$P(X_s=x_s | X_r=x_r, r \neq s) = P(X_s=x_s | X_r=x_r, r \in \mathcal{G}_s)$$

$$\forall s \in S, \omega = (x_{s_1}, x_{s_2}, \dots, x_{s_N}) \in \Omega$$

(4.6)

L'insieme di tutte le probabilità condizionali  $P(X_s=x_s | X_r=x_r, r \neq s)$  è detto caratteristica locale del MRF e determina univocamente la probabilità congiunta  $P(X=\omega)$  secondo la seguente relazione:

$$P(X_s=x_s | X_r=x_r, r \neq s) = \frac{P(X=\omega)}{\sum_{x_s \in \Lambda} P(X=\omega)} \quad (4.7)$$

Dato l'insieme dei siti  $S$  e il sistema di intorno  $\mathcal{G}$  su di esso definito, si può definire un altro insieme, l'insieme delle clique  $\mathcal{C}$ , costituito da tutti i sottoinsiemi  $C$  di  $S$  le cui coppie di elementi (siti) distinti stanno uno nell'intorno dell'altro. In altre parole, una clique è costituita da siti interagenti a due a due, nel senso che la probabilità condizionale dell'uno dipende dall'altro e viceversa.

Si definisce distribuzione di Gibbs ogni misura di probabilità  $\pi$ , relativa a  $\Omega$  e alla coppia  $(S, \mathcal{G})$  con insieme di clique associato  $\mathcal{C}$ , rappresentabile come:

$$\pi(\omega) = \frac{1}{Z} \exp \left[ -\frac{U(\omega)}{T} \right] \quad (4.8)$$

dove

$$Z \equiv \sum_{\omega} \exp \left[ -\frac{U(\omega)}{T} \right] \quad (4.9)$$

è chiamata funzione di partizione ed ha il significato di una costante normalizzante. La costante  $T$ , che verrà nel seguito chiamata temperatura, regola il grado di "peaking" di  $\pi$ . Valori piccoli di  $T$  aumentano le mode di  $\pi$ , rendendo così più facile trovarle mediante campionamento. E' questo il principio su cui si basa l'annealing simulato.

La quantità

$$U(\omega) = \sum_{C \in \mathcal{C}} V_C(\omega) \quad (4.10)$$



è detta funzione energia e l'insieme

$$\{V_C, C \in \mathcal{C}\} \quad (4.11)$$

è detto potenziale. Deve essere precisato che le funzioni  $V_C(\omega)$  dipendono solo da quelle coordinate  $x_s$  di  $\omega$  per cui  $s$  appartiene alla clique  $C$ , cioè sono funzioni che attribuiscono un certo valore di "potenziale" alle possibili configurazioni di gruppi di variabili aleatorie che interagiscono fra loro. In pratica, in questa formulazione, la probabilità congiunta del campo risulta essere il prodotto delle probabilità congiunte delle variabili casuali associate alle varie clique che compongono il sistema dei siti. Risulta quindi evidente come le informazioni sul comportamento locale del campo possano, in questo modo, essere inserite direttamente nella probabilità congiunta, anziché dar luogo alle singole probabilità condizionali.

Vale il seguente importante

#### TEOREMA [28]

Sia  $\mathcal{G}$  un sistema di intornoi su un insieme di siti  $S$ . Allora  $X$  è un MRF rispetto a  $\mathcal{G}$  se e solo se  $P(X=\omega)$  è una distribuzione di Gibbs rispetto a  $\mathcal{G}$ .

Il fatto che la densità di probabilità congiunta di un MRF sia esprimibile mediante una distribuzione di Gibbs rende più facile caratterizzare il MRF stesso. Basta infatti specificare il potenziale, che può essere facile, anziché la caratteristica locale, che è senz'altro più difficile. Allora  $U$  può essere scelta in accordo al comportamento locale del MRF, desiderato o noto, e, nel caso di processi immagine, questo è fattibile almeno a livello del processo di intensità  $F$ . Quello che viene modellato, attraverso la funzione energia, è l'interrelazione fra elementi di una clique, ma è facile dimostrare come da questo discenda immediatamente la dipendenza di una variabile casuale del campo da tutti gli elementi

che stanno nel suo intorno. Infatti, per una densità di probabilità Gibbsiana, le probabilità condizionali possono essere calcolate mediante l'integrazione (4.7), e risultano essere date da:

$$\begin{aligned}
 P(X_s=x_s | X_r=x_r, r \neq s) &= P(X_s=x_s | X_r=x_r, r \in \mathcal{A}_s) = \\
 &= \frac{1}{Z_s} \exp \left[ -\frac{1}{T} \sum_{C: s \in C} V_C(\omega) \right]
 \end{aligned} \tag{4.12}$$

con

$$Z_s \equiv \sum_{x \in \Lambda} \exp \left[ -\frac{1}{T} \sum_{C: s \in C} V_C(\omega^x) \right] \tag{4.13}$$

dove  $\omega^x$  è la configurazione che ha  $x$  al sito  $s$ . La condizione caratterizzante un MRF, cioè che la probabilità condizionale di  $X_s$  è legata alle sole altre variabili aleatorie ad esso adiacenti secondo il sistema di interni prescelto, discende immediatamente dalla (4.12) osservando che

$$\bigcup_{C: s \in C} C = \mathcal{A}_s \tag{4.14}$$

## 5. Modelli MRF per immagini

Un modello per immagini a MRF si presta bene a rappresentare, oltre al processo di intensità, anche altri attributi, tipicamente geometrici, che caratterizzano le immagini del mondo reale. Come esempi possono essere citati gli edge, cioè i contorni degli oggetti contenuti nell'immagine, e le componenti connesse, cioè le aree dell'immagine a valori in intervalli prefissati, che identificano gli oggetti, o parti di essi [13,32].

Nel seguito, con riferimento a [12], l'immagine originale è vista come un processo  $X$  decomponibile in una coppia  $(F,L)$  dove  $F$  è la matrice osservabile delle intensità dei pixel, precedentemente

definita, e  $L$  è una nuova matrice di variabili casuali inosservabili associate agli elementi di edge, idealmente collocati negli spazi interpixel. Si parlerà di processo di intensità per  $F$  e di processo di linea per  $L$ . Entrambi questi processi sono marginali rispetto al processo globale  $X=(F,L)$ .

Occorre osservare che ciascuno dei due processi  $F$  ed  $L$  avrà il suo proprio spazio dei valori, spazio delle configurazioni, sistema di intorni e di clique associate. Nel caso più semplice  $L$  sarà visto come un processo binario, cioè a due soli valori, 0 (elemento di linea non attivo) e 1 (elemento di linea attivo). In questo modo risultano ammesse due sole orientazioni, verticale, per l'elemento di linea posto fra due pixel adiacenti sulla stessa riga dell'immagine, ed orizzontale, per l'elemento di linea posto fra due pixel adiacenti sulla stessa colonna. Volendo includere nel modello la possibilità di avere elementi di linea variamente orientati, occorrerà aumentare il numero dei valori (stati), stabilendo una corrispondenza fra essi e le orientazioni di linea che si vogliono considerare.

Modellare il processo immagine  $X=(F,L)$  come un MRF significa assegnare per esso un sistema di intorni  $\mathcal{G}$  e una funzione energia  $U$  che esprima le dipendenze locali fra pixel e/o elementi di linea. Come prima cosa verranno analizzati alcuni tipici sistemi di intorni utilizzati in letteratura.

### Sistemi di intorni per siti di intensità

Si consideri il solo processo di intensità  $F$ . In questo caso, l'insieme dei siti  $S$  sarà dato da ogni ordinamento del reticolo cartesiano  $m \times m$  di punti o pixel su cui l'immagine  $F$  è definita, ad esempio  $S = Z_m = \{(i,j), i=1,2,\dots,m ; j=1,2,\dots,m\}$  con  $N=m^2$ .

Su  $S$  può essere definito un sistema di intorni omogeneo del tipo:

$$\mathcal{G} = \{ \mathcal{F}_{ij} : (i,j) \in Z_m \}$$

$$\mathcal{F}_{ij} = \{ (k,l) \in Z_m : 0 < (k-i)^2 + (l-j)^2 \leq r \} \quad (5.1)$$

dove  $r$  è una costante.

Per  $r=1$  (primo ordine o nearest-neighbor) la configurazione di intorno per il pixel identificato col simbolo  $\bullet$  è data in Fig.1.

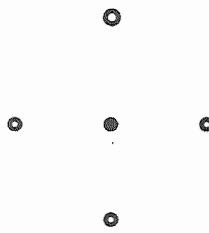


Fig.1 - intorno del primo ordine

Per  $r=2$  (secondo ordine) la configurazione di intorno per lo stesso pixel è data in Fig.2.

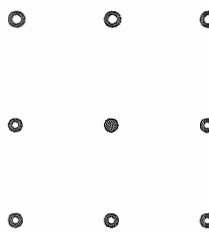


Fig.2 - intorno del secondo ordine

Si nota che i siti di bordo devono avere un intorno costituito da un numero inferiore di pixel. Questa condizione è detta condizione di "free boundary" e, in elaborazione di immagini, risulta più naturale dell'assumere un reticolo toroidale o periodico in genere.

L'insieme delle cliques dipende dal sistema di intorni scelto ed è da esso univocamente determinato. Le configurazioni di clique

per  $r=1$  sono illustrate in Fig.3 mentre quelle per  $r=2$  sono mostrate in Fig.4.

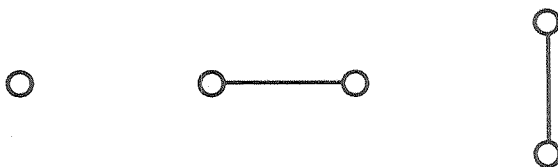


Fig. 3 - sistema di clique per intorni del primo ordine

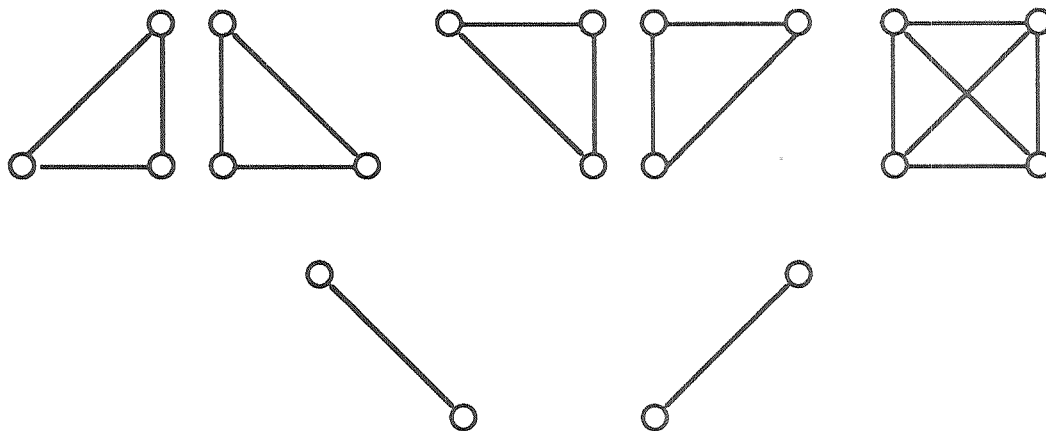


Fig.4 - sistema di clique per intorni del secondo ordine

Sistema di intorni per siti di linea

Considerando il solo processo di linea  $L$ , l'insieme dei siti  $S=D_m$  che risulta naturale associargli è quello costituito dai punti localizzati a metà fra ogni coppia verticale ed orizzontale di pixel. Questi siti rappresentano le possibili locazioni degli elementi di edge di oggetti nell'immagine. Allora un esempio di sistema di intorni a 6 punti definibile su  $D_m$  ed identificato come  $L_d$  ammette le due configurazioni riportate in Fig.5.

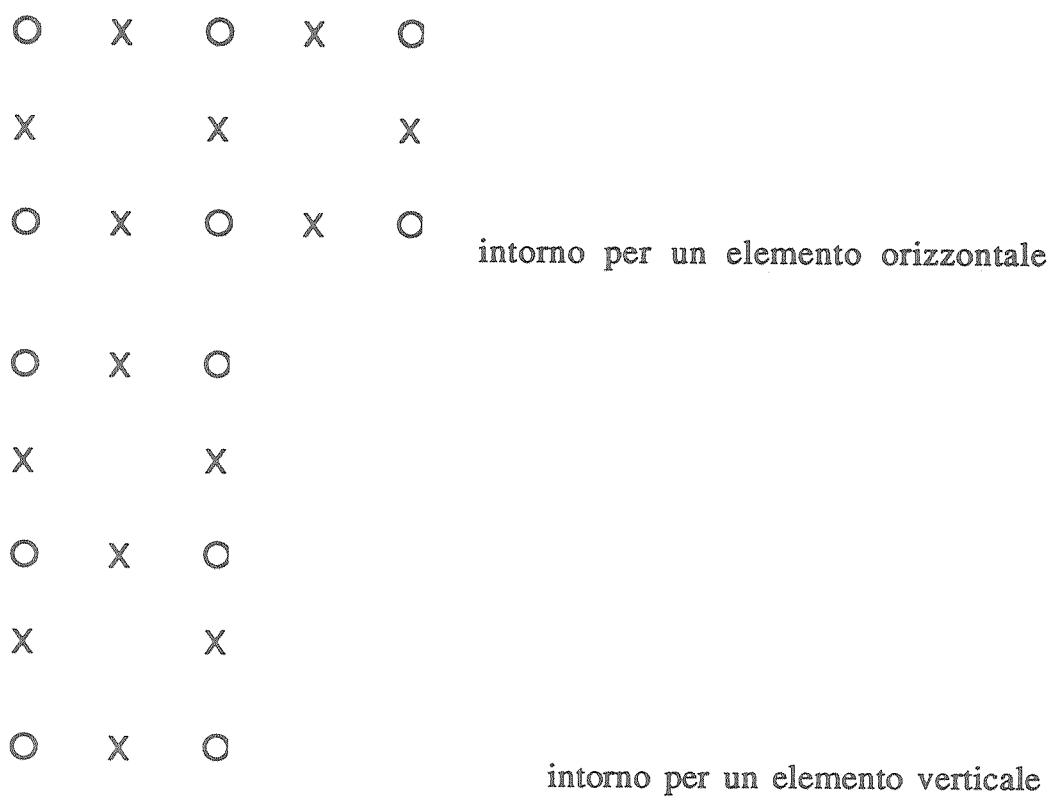


Fig.5 - intorni per elementi di linea

Tre possibili configurazioni di clique per questo sistema di intorni sono mostrate in Fig.6.

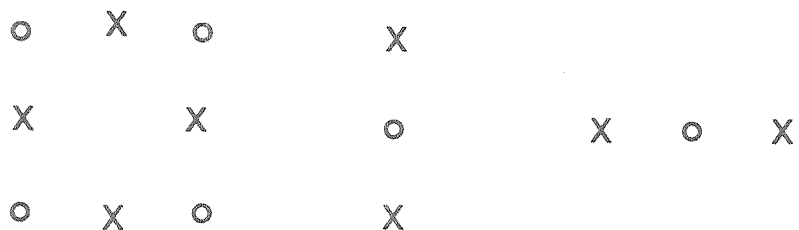


Fig.6 - esempi di clique per il sistema di intorni di Fig.5

Sistema di intorni misto

Considerando il campo completo  $X=(F,L)$ , un naturale insieme dei siti  $S$  per  $X$  sarà dato da  $S=Z_m \cup D_m$ . Un possibile sistema di intorni su  $S$  è definito dall'unione di  $\mathcal{F}_1$  (nearest-neighbor) per il processo di intensità  $F$  e di  $\mathcal{L}_d$  per il processo di linea  $L$ . L'insieme di clique associate, in questo caso, sarà dato dall'unione delle clique di intensità e di quelle di linea.

Si consideri adesso il campo  $X=(F,L)$  con insieme dei siti  $S=Z_m \cup D_m$ , e con associato un certo sistema di intorni  $\mathcal{G}$ . Lo spazio delle configurazioni sia l'insieme delle coppie  $\omega=(f,l)$  dove  $f$  assume valori nel range di valori di intensità (livelli di grigio) ammessi e  $l$  nell'insieme degli stati di linea (orientazioni) definiti. Allora la densità di probabilità a priori gibbsiana per  $X$  sarà data da:

$$P(F=f, L=l) = \frac{1}{Z} \exp \left[ -\frac{U(f,l)}{T} \right] \quad (5.2)$$

dove nel seguito si porrà  $T=1$  per semplicità, e con

$$U(f,l) = \sum_C V_C(f,l) \quad (5.3)$$

la funzione energia.

Adesso i termini di potenziale potranno essere definiti su clique omogenee (costituite di soli siti di intensità o di soli siti di linea) oppure miste (costituite sia da siti di linea che da siti di intensità, fra essi adiacenti). In generale la (5.3) potrà quindi essere riscritta come:

$$U(f,l) = U_1(f) + U_2(l) + U_3(f,l) \quad (5.4)$$

I termini di potenziale associati alle sole variabili di intensità avranno la funzione di modellare il comportamento locale del processo di intensità dell'immagine, mentre i termini di potenziale nelle sole variabili di linea modelleranno il comportamento locale

del processo di linea, attribuendo valori di "potenziale", e quindi di probabilità, a particolari configurazioni di linea, quali terminazioni, intersezioni, parallelismi, continuazioni. In questo modo potranno essere incorporate nella probabilità a priori per l'immagine informazioni che si hanno sulla struttura geometrica dei contorni degli oggetti in essa contenuti.

I termini di potenziale nelle variabili di intensità e di linea sono i più interessanti, in quanto consentono di modellare le interazioni reciproche fra elementi di linea ed elementi di intensità, mediante l'attribuzione di valori di probabilità a configurazioni locali di tali elementi.

Esempi di tipici termini di potenziale per immagini verranno dati nel paragrafo 9.

## 6. La probabilità a posteriori

Si è visto che, nell'assunzione di un'immagine modellabile come un MRF, è possibile definire una sua densità di probabilità a priori  $P(X=\omega)$ , in termini di una distribuzione di Gibbs. Di particolare interesse è la definizione del sistema di intorno che si intende adottare, in quanto esso esprime la geometria delle interrelazioni fra pixel adiacenti e fra elementi di linea adiacenti. La definizione di una funzione energia consente poi di dare un valore alla probabilità globale che si verifichi una certa configurazione (probabilità congiunta), e, più in particolare, alla probabilità che un dato sito assuma un certo valore, assegnati i valori dei siti adiacenti (probabilità condizionale).

Di particolare interesse è anche l'esame di come sia possibile derivare la probabilità a posteriori di un'immagine modellata come un MRF che sia stata degradata, cioè la probabilità che a un certo insieme di dati o di misure corrisponda una certa configurazione per l'immagine da stimare. La massimizzazione di tale probabilità a posteriori è il principio su cui si basano le stime MAP.

Per semplicità si supponga che nel modello di degradazione dato dall'eq.(2.1) (ancora relativo al solo processo di intensità, cioè



il processo di linea non viene trasformato),  $H$  rappresenti un operatore lineare ed invariante per traslazioni. Si supponga inoltre che  $N$  sia un processo di rumore bianco gaussiano con media  $\mu$  e varianza  $\sigma^2$ , indipendente da  $X$ .

Quello che si vuole ricavare è un'espressione per  $P(X=\omega \mid G=g)$ . Si vedrà che tale probabilità a posteriori è ancora esprimibile come una distribuzione di Gibbs su un sistema di intorni allargato rispetto a quello considerato per  $X$ .

Si assuma che l'insieme dei siti sia dato dal tipico e già definito insieme  $S=Z_m \cup D_m$ , con associato un certo sistema di intorni  $\mathcal{G}$ . Lo spazio delle configurazioni sia l'insieme delle coppie  $\omega=(f,l)$  dove  $f$  assume valori nel range di livelli di grigio ammessi e  $l$  nell'insieme degli stati di linea definiti. Allora si ricorda che la densità di probabilità a priori per  $X$  sarà data da:

$$P(F=f, L=l) = \frac{1}{Z} \exp \left[ -\frac{U(f,l)}{T} \right] \quad (6.1)$$

dove nel seguito porremo  $T=1$  per semplicità, e con

$$U(f,l) = \sum_C V_C(f,l) \quad (6.2)$$

la funzione energia.

Sia  $\mathcal{H}_s$ ,  $s$  in  $Z_m$ , l'insieme dei pixel che concorrono alla formazione del pixel  $s$  nell'immagine distorta  $H(F)$ . Per esempio, se  $H$  corrisponde ad un filtro  $3 \times 3$ ,  $\mathcal{H}_s$  sarà il quadrato  $3 \times 3$  centrato in  $s$ .

Allora l'insieme

$$\{ \mathcal{H}_s \setminus \{s\} \mid s \in Z_m \} \quad (6.3)$$

è un sistema di intorni su  $Z_m$  e anche l'insieme

$$\begin{aligned} & \{ \mathcal{H}_s^2 \setminus \{s\} \mid s \in Z_m \} \text{ con} \\ & \mathcal{H}_s^2 = \cup \mathcal{H}_r \quad r \in \mathcal{H}_s \quad \forall s \in Z_m \end{aligned} \quad (6.4)$$

lo è. Si vedrà che il sistema di intorni su  $S$  definito come

$$\mathcal{G}^P = \{ \mathcal{G}_s^P, s \in S \} \quad \text{dove}$$

$$\mathcal{G}_s^P = \begin{cases} \mathcal{G}_s & s \in D_m \\ \mathcal{G}_s \cup \mathcal{H}_s^2 \setminus \{s\} & s \in Z_m \end{cases} \quad (6.5)$$

costituisce il sistema di intorni per la densità di probabilità a posteriori.

#### TEOREMA [12]

Per ogni  $g$  fissato,  $P(X=\omega \mid G=g)$  è una distribuzione di Gibbs su  $\{S, \mathcal{G}^P\}$  con funzione energia data da:

$$U^P(f, l) = U(f, l) + \|g - Hf - \mu\|^2 / 2\sigma^2 \quad (6.6)$$

dove  $\mu$  è il vettore di  $R^N$  che ha tutte le componenti uguali a  $\mu$ .

#### Dimostrazione

Per il teorema di Bayes vale:

$$P(X=\omega \mid G=g) = \frac{P(G=g \mid X=\omega) P(X=\omega)}{P(G=g)}$$

Dato che  $P(G=g)$  è una costante e

$$P(X=\omega) = \frac{1}{Z} e^{-U(\omega)}$$

rimane da calcolare  $P(\mathbf{G}=\mathbf{g} \mid \mathbf{X}=\omega)$ . Si trova che:

$$\begin{aligned} P(\mathbf{G}=\mathbf{g} \mid \mathbf{X}=\omega) &= P(\mathbf{H}\mathbf{F} + \mathbf{N}=\mathbf{g} \mid \mathbf{F}=\mathbf{f}, \mathbf{L}=\mathbf{l}) = \\ &= P(\mathbf{N}=\mathbf{g}-\mathbf{H}\mathbf{f} \mid \mathbf{F}=\mathbf{f}, \mathbf{L}=\mathbf{l}) = P(\mathbf{N}=\mathbf{g}-\mathbf{H}\mathbf{f}) \end{aligned} \quad (6.7)$$

data l'indipendenza di  $\mathbf{N}$  da  $\mathbf{F}$  e  $\mathbf{L}$ . Nell'assunzione di  $\mathbf{N}$  Gaussiano si ha:

$$P(\mathbf{N}=\mathbf{g}-\mathbf{H}\mathbf{f}) = (2\pi \sigma^2)^{-\frac{N}{2}} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{g}-\mathbf{H}\mathbf{f}-\boldsymbol{\mu}\|^2\right) \quad (6.8)$$

L'espressione (6.8) era già stata ricavata, in un contesto più generale ed indipendentemente dalla formulazione in termini di distribuzioni di Gibbs, nel paragrafo 2, nel caso di rumore a media nulla, e per  $\mathbf{H}$  generale. Data la (6.8) si può scrivere:

$$P(\mathbf{X}=\omega \mid \mathbf{G}=\mathbf{g}) = \frac{1}{Z^P} e^{-U^P(\omega)} \quad (6.9)$$

con  $U^P$  definito nell'enunciato del teorema, e  $Z^P$  la costante normalizzante che ora dipenderà da  $\mathbf{g}$ .

Rimane da determinare il sistema di intorno relativo ad  $U^P$ . Innanzi tutto è intuitivo osservare che i siti di linea dovranno mantenere lo stesso sistema di intorno su cui è definita la probabilità a priori. Viceversa l'intorno  $\mathcal{G}_s$  di un sito pixel in  $Z_m$  dovrebbe risultare aumentato per tenere conto del processo di degradazione. Il calcolo delle caratteristiche locali per la probabilità a posteriori chiarirà questi concetti intuitivi.

Per il processo di linea  $\mathbf{L}$  si avrà:

$$P(L_s=l_s | L_r=l_r, r \neq s, r \in D_m, F=f, G=g) \equiv$$

$$\equiv \frac{e^{-U^P(f,l)}}{\sum_{l_s} e^{-U^P(f,l)}} = \frac{e^{-U(f,l)}}{\sum_{l_s} e^{-U(f,l)}} \quad (6.10)$$

Quindi  $\mathcal{G}_s^P = \mathcal{G}_s \quad \forall s \in D_m$

Per il processo di intensità  $F$  vale invece:

$$P(F_s=f_s | F_r=f_r, r \neq s, r \in Z_m, F=f, G=g) \equiv$$

$$\equiv \frac{e^{-U^P(f,l)}}{\sum_{f_s} e^{-U^P(f,l)}} \quad (6.11)$$

Ora  $U^P(f,l)$  può essere decomposto nel seguente modo:

$$\begin{aligned} U^P(f,l) &= U(f,l) + \sum_{r \in Z_m} (g_r - (Hf)_r - \mu)^2 / 2\sigma^2 = \\ &= \sum_{C:s \in C} V_C(f,l) + \sum_{C:s \in C} V_C(f,l) + \\ &+ \sum_{r:s \in \mathcal{H}_r} (g_r - H(f_t, t \in \mathcal{H}_r) - \mu)^2 / 2\sigma^2 + \\ &+ \sum_{r:s \in \mathcal{H}_r} (g_r - H(f_t, t \in \mathcal{H}_r) - \mu)^2 / 2\sigma^2 \end{aligned} \quad (6.12)$$

I termini 2° e 4° della (6.12), che non contengono  $f_s$ , risulteranno così semplificati nel rapporto (6.11), che dipenderà

allora solo dai termini 1° e 3°. In particolare il 1° termine dipende dalle coordinate di  $(f, l)$  per siti in  $\mathcal{G}_s$ , e il 3° termine dipende dalle coordinate di  $f$  per siti che stanno in:

$$\bigcup_{r: s \in \mathcal{H}_r} \mathcal{H}_r = \bigcup_{r \in \mathcal{H}_s} \mathcal{H}_r \equiv \mathcal{H}_s^2 \quad (6.13)$$

Quindi

$$\mathcal{G}_s^P = \mathcal{G}_s \cup \mathcal{H}_s^2 \setminus \{s\} \quad \forall s \in Z_m \quad (6.14)$$

come affermato nel Teorema.

## 7. Algoritmi di Rilassamento Stocastico

Una volta calcolata la densità di probabilità a posteriori  $P(\mathbf{X}=\omega|\mathbf{G}=\mathbf{g})$  per  $\mathbf{g}$  assegnato, lo scopo della stima MAP è quello di calcolare i valori di  $\omega$  che la massimizzano, cioè che minimizzano:

$$U(f, l) + \|\mathbf{g} - \mathbf{H}f - \boldsymbol{\mu}\|^2 / 2\sigma^2 \quad \text{con } (f, l) \text{ in } \Omega \quad (7.1)$$

Questo problema computazionale è in generale molto difficile da risolvere, data la complessità della funzione obiettivo e la dimensione molto elevata di  $\Omega$ . Tipicamente si ha infatti che con l'introduzione del processo di linea, il funzionale costo (7.1) risulta non convesso, e potrà quindi presentare vari minimi locali, possibilmente anche molto lontani dal minimo globale che si vuole stimare [9].

Si è già detto come, in questi casi, un classico metodo di descent deterministico tenderà a bloccarsi in corrispondenza del minimo locale più vicino al punto di partenza.

Fra gli algoritmi proposti per superare questo ostacolo, molto popolari sono gli algoritmi di rilassamento stocastico. La filosofia generale di tali algoritmi è che possono essere accettati (in probabilità) anche cambi nelle configurazioni del processo da stimare che aumentano il valore del funzionale costo. Questo

avviene proprio allo scopo di saltare via da eventuali minimi locali nei quali si possa essere caduti.

Chiamando per semplicità  $E(\omega)$  il funzionale che deve essere minimizzato, ad esempio il funzionale MAP di eq. (7.1), si esaminerà di seguito il funzionamento di alcuni fra i più noti algoritmi di rilassamento stocastico.

Il metodo Monte Carlo [12], forse il più conosciuto, consiste nella generazione casuale (ad esempio con legge uniforme) di una sequenza di configurazioni  $\omega$ , nella valutazione del funzionale  $E(\omega)$  su tale sequenza e nella scelta di quella configurazione a cui corrisponde il valore più basso per  $E$ . Questo metodo, anche se molto attraente per la sua semplicità, ha l'inconveniente di essere in generale molto costoso da un punto di vista computazionale.

Una versione più sofisticata di questo metodo è rappresentata dall'algoritmo di Metropolis [18], che fu ideato per studiare le proprietà di equilibrio di sistemi costituiti da un grande numero di componenti essenzialmente identiche e interagenti fra loro, come ad esempio molecole in un gas.

Sia  $\Omega$  l'insieme di tutte le possibili configurazioni del sistema, e sia  $E(\omega)$  la funzione energia che lo caratterizza. Allora, se il sistema è in equilibrio termico con l'ambiente circostante, la probabilità di realizzazione di una configurazione  $\omega$  in  $\Omega$  sarà data da:

$$\pi(\omega) = \frac{e^{-\beta E(\omega)}}{\sum_{\eta} e^{-\beta E(\eta)}} \quad \omega \in \Omega \quad (7.2)$$

dove  $\beta=1/kT$  ( $K$  costante di Boltzmann,  $T$  temperatura assoluta) e  $\pi$  è detta Fattore di Boltzmann. Dovendo minimizzare l'energia  $E(\omega)$ , l'idea di base dell'algoritmo di Metropolis è quella di scegliere i campioni da  $\pi$  invece che uniformemente come nel metodo Monte Carlo. Chiaramente, la legge di probabilità sopra descritta,

indipendentemente dal valore di  $\beta$ , risulta massima laddove  $E$  risulta minima, e, sempre a meno di  $\beta$ , ha la stessa forma della probabilità a posteriori nelle stime MAP con l'utilizzo di modelli a MRF. Questo evidenzia la diretta applicabilità dell'algoritmo al problema in esame in questo lavoro. L'algoritmo funziona in pratica nel seguente modo:

1. Dato lo stato del sistema al tempo  $t$ , detto  $X(t)$ , si sceglie un'altra configurazione  $\eta$  a caso (ad esempio con legge uniforme).
2. Si valuta il cambio nell'energia dato da  $\Delta E = E(\eta) - E(X(t))$ .
3. Se  $\Delta E \leq 0$  (l'energia è diminuita) lo spostamento in  $\eta$  è permesso e si pone  $X(t+1) = \eta$ .
4. Se  $\Delta E > 0$  (l'energia è aumentata) la transizione è fatta con probabilità data da  $e^{-\beta \Delta E}$ . In pratica si opera come segue: si estrae un numero  $\xi$ ,  $0 \leq \xi \leq 1$  con distribuzione uniforme e si pone:

$$X(t+1) = \eta \quad \text{se } \xi \leq e^{-\beta \Delta E}$$

$$X(t+1) = X(t) \quad \text{se } \xi > e^{-\beta \Delta E}$$

Questo significa che sono accettate soltanto transizioni in accordo con la (7.2).

Possono essere fatte le seguenti osservazioni:

- Per  $T$  fissato, la probabilità di accettazione di  $\Delta E$  diminuisce al crescere di  $\Delta E$  stesso, cioè più è grande l'aumento di energia prodotto dalla transizione e più è scarsa la probabilità che si ha di accettare tale transizione.
- Per  $\Delta E$  fissato, la sua probabilità di accettazione è tanto più alta quanto più è grande  $T$ .

Allora  $T$  sarà scelto sulla base dell'ampiezza dei  $\Delta E$  che si vogliono accettare con alta probabilità.

L'essenza dell'algoritmo di Metropolis, così come quella di ogni schema di rilassamento stocastico, è in generale che cambi nella configurazione ( $\omega \rightarrow \eta$ ) che aumentano l'energia (cioè che diminuiscono la probabilità  $\pi$ ) sono permessi, proprio allo scopo di saltare via dai minimi locali, mentre gli algoritmi deterministici consentono solo transizioni verso stati che diminuiscono l'energia, e invariabilmente si bloccano in corrispondenza di minimi locali.

Quindi i metodi di rilassamento stocastico sono in generale più efficienti dei metodi deterministici nelle stime di tipo MAP. Un ulteriore miglioramento a tali metodi può essere apportato introducendo nello schema computazionale anche un procedimento di annealing simulato [15].

Questo consiste nell'introdurre in  $\pi$  un parametro  $T$  variabile, detto Temperatura, che ne regola il grado di "peaking", e che viene fatto gradualmente abbassare da un certo valore iniziale. Allora per alti valori di  $T$ , l'applicazione di un metodo di rilassamento stocastico (per esempio l'algoritmo di Metropolis) porterà a transizioni che, con probabilità alta, possono aumentare il valore dell'energia. Idealmente, per  $T$  infinito l'algoritmo si comporterebbe come il metodo Monte Carlo, con la scelta della nuova configurazione su una base puramente casuale, in accordo con la distribuzione  $e^{-\beta \Delta E}$ , praticamente uniforme. Questo consente di superare anche i minimi locali più profondi. Per bassi valori di  $T$  l'algoritmo tenderà a consentire solo transizioni che diminuiscono il valore dell'energia.

L'idea dell'annealing simulato è quindi quella di iniziare l'algoritmo con un valore alto di  $T$ , per esempio scelto in modo tale che, con alta probabilità, consenta di saltare fuori anche dai minimi locali più profondi. Via via che si procede, la temperatura viene gradualmente abbassata, in modo da ridurre sempre di più l'ampiezza delle oscillazioni permesse. Idealmente, una volta



arrivati nella valle del minimo globale, il valore di temperatura dovrebbe essere tanto basso da inibire transizioni verso configurazioni che alzano anche di poco l'energia, in modo tale da essere sicuri di non poter più uscire da tale valle e, anzi, di spostarsi sempre più verso il minimo globale.

L'efficienza del metodo consiste nell'individuare un opportuno schema di annealing, cioè una sequenza decrescente di temperature  $\{T_n\}$  e di tempi di sosta associati ad ogni valore di tale sequenza, in modo tale da avere la ragionevole certezza che tutto il metodo converga al massimo globale di  $\pi$ .

In Kirkpatrick et al. [15] lo schema di annealing è determinato ad hoc, per analogia con il comportamento, ad una temperatura finita, di un sistema in equilibrio con molti gradi di libertà.

In [12], invece, viene dimostrata l'esistenza di uno schema di annealing che garantisce la convergenza al massimo globale della probabilità a posteriori.

La novità più significativa di tale schema consiste nel fatto che la generazione dei campioni, in accordo ad una assegnata probabilità a posteriori, viene ottenuta mediante un processo iterativo di rimpiazzamento locale dei singoli elementi di intensità e di linea. Il rimpiazzamento viene eseguito sulla base della probabilità condizionale a posteriori di un elemento dati i valori assunti dagli elementi limitrofi e può essere eseguito in parallelo contemporaneamente su tutti gli elementi.

L'equivalenza fra MRF e distribuzioni di Gibbs viene sfruttata per il calcolo delle probabilità condizionali, secondo quanto dimostrato nei paragrafi 4 e 6.

## 8. Campionatore di Gibbs

Si riporta di seguito una descrizione informale dell'algoritmo di rilassamento stocastico per la stima MAP proposto in [12]; tale algoritmo è basato sullo schema sopra citato di generazione di campioni e viene detto Campionatore di Gibbs.

In [12] è dimostrato che tale algoritmo, combinato con uno schema di annealing, produce una successione di configurazioni  $X_T(t)$ , dipendenti dal parametro  $T$  (temperatura), che converge alla soluzione. Le linee di sviluppo dell'algoritmo sono le seguenti.

1. Il procedimento inizia con valori di temperatura alti che, potendo portare a cambi nell'immagine che decrescono la distribuzione a posteriori, consentono di superare i vari massimi locali. Al procedere del rilassamento, la temperatura è gradualmente abbassata, in modo che il processo si comporta sempre di più come un algoritmo iterativo deterministico che accresce il valore della distribuzione a posteriori.
2. Un "Teorema di Annealing" fornisce una legge per abbassare la temperatura che garantisce la convergenza dell'algoritmo al massimo globale della distribuzione a posteriori.
3. Nella configurazione corrente  $X_T(t)$  dell'immagine viene fatto un rimpiazzamento locale di ciascun elemento (valore del singolo pixel o orientazione e localizzazione del singolo elemento di linea) sulla base dei valori dei siti adiacenti. Tale rimpiazzamento viene eseguito estraendo un campione dalla locale distribuzione di probabilità condizionale a posteriori. Questo processo, nella sua applicazione all'intera immagine, può essere eseguito contemporaneamente su tutti i punti e quindi risulta altamente parallelo.

Poichè le distribuzioni condizionali locali dipendono dalla temperatura, a "basse" temperature esse si concentrano su stati che aumentano il valore della distribuzione a posteriori congiunta; per "alte" temperature le densità condizionali sono approssimativamente uniformi e possono portare anche a diminuzioni nel valore della distribuzione a posteriori.

Il procedimento di rimpiazzamento dei siti viene descritto più in dettaglio nel seguito, con riferimento ad una architettura parallela, per una prefissata temperatura.

Si immagini di avere un semplice processore posizionato in ogni sito. Il processore al sito  $s$  è connesso ai processori dei siti in  $\mathcal{G}_s$ . Il numero di siti e quindi di processori per l'intero processo immagine  $X$  è molto grande, tuttavia le dimensioni degli intorno, e quindi il numero di connessioni a un singolo processore, può essere molto modesto, pur tenendo conto dei pixel, delle linee e dei legami misti fra pixel e linee. Si supponga per il momento che esista un qualche tipo di sincronizzazione tra i processori, e che sia possibile individuare una sequenza  $\{n_t, t=1,2,3,\dots\}$  secondo cui i siti sono visitati per il rimpiazzamento.

Una scelta naturale per  $\{n_t\}$  sarà per esempio quella corrispondente ad una scansione raster (dall'alto verso il basso e da sinistra a destra) del reticolo immagine. In questo caso, ovviamente, può essere utilizzato un solo processore, da spostare via via nei vari siti, rendendo così seriale il procedimento.

Ogni processore è programmato per seguire lo stesso algoritmo. Al tempo  $t$ , che individua il sito  $s_{n_t}$ , un campione  $x$  è tratto dalla probabilità condizionale di  $X_{n_t}$ , dati gli stati osservati dei siti adiacenti  $X_r(t-1)$ ,  $r$  in  $\mathcal{G}_{n_t}$ . Tale campione  $x$  sarà un elemento dello spazio degli stati  $\Lambda$ . La nuova configurazione  $X(t)$  è data da:

$$\begin{aligned} X_{n_t}(t) &= x \\ X_s(t) &= X_s(t-1) \quad \text{per } s \neq s_{n_t} \end{aligned} \quad (8.1)$$

Data la configurazione  $X(0)$ , arbitraria, si ottiene così una sequenza  $X(1), X(2), X(3), \dots$  di configurazioni con proprietà di convergenza dimostrabili teoricamente. Il limite ottenuto per la sequenza non dipenderà da  $X(0)$ .

E' fondamentale osservare che il risultato stabilito sopra è valido per qualunque sequenza  $\{n_t\}$ , e quindi è consentito uno schema asincrono in cui ogni processo può essere guidato dal suo proprio clock. In questo modo il procedimento risulta altamente parallelo in quanto può essere realizzato mediante semplici unità che operano in modo sostanzialmente indipendente. L'unica forma di dipendenza fra processori è data dalle connessioni che devono esistere coi processori dei siti adiacenti. Ne deriva che il tempo richiesto per un completo aggiornamento dell'intero sistema è indipendente dal numero dei siti.

Con  $\pi(\omega)$  si indichi ancora una volta la probabilità a posteriori di avere  $X=\omega$  dato  $G=g$ , cioè:

$$\pi(\omega) = P(X=\omega \mid G=g) = 1/Z^P \exp [-UP(\omega)] \quad (8.2)$$

Vale il seguente

#### TEOREMA (di rilassamento)[12]

Si assuma che, per ogni  $s$  in  $S$ , la sequenza  $\{n_t\}$  contenga  $s$  infinite volte. Allora per ogni configurazione iniziale  $\eta$  in  $\Omega$  e per ogni  $\omega$  in  $\Omega$  si ha:

$$\lim_{t \rightarrow \infty} P ( X(t)=\omega \mid X(0)=\eta ) = \pi (\omega)$$

Come l'algoritmo di Metropolis, anche il campionatore di Gibbs produce una catena di Markov  $\{X(t), t=1,1,2,\dots\}$  con  $\pi$  come distribuzione di equilibrio.

In pratica, il significato di questo teorema è che il campionatore di Gibbs, a temperatura costante, asintoticamente produce, per ogni configurazione iniziale del campo, una configurazione che è in accordo alla distribuzione densità di probabilità a posteriori  $\pi$  che deve essere massimizzata, come richiesto dai metodi di rilassamento stocastico di tipo Metropolis.

Ponendo  $\pi = P(X = \omega)$ , la densità di probabilità a priori per il processo immagine, il campionatore di Gibbs a temperatura  $T=1$  può essere utilizzato per generare un campione da tale probabilità, cioè un'immagine che appartiene alla classe che soddisfa certe informazioni a priori, formulate in termini di densità di probabilità.

Si supponga ora di far dipendere  $\pi$  da un parametro  $T(t)$  che verrà indicato come la temperatura al tempo  $t$ . Allora si definisca:

$$\pi_{T(t)}(\omega) = \frac{1}{Z_{T(t)}^P} e^{-U^P(\omega)/T(t)} \quad (8.3)$$

L'introduzione di tale parametro nello schema computazionale definito dal campionatore di Gibbs dà luogo ad una procedura di annealing che genera un processo diverso  $\{X(t), t=1,2,\dots\}$  tale che:

$$\begin{aligned} P(X_s(t) = x_s, s \in S) &= \\ &= \pi_{T(t)}(X_{n_t} = x_{n_t} | X_s = x_s, s \neq n_t) P(X_s(t-1) = x_s, s \neq n_t) \end{aligned} \quad (8.4)$$

Sia

$$\Omega_0 = \{ \omega \in \Omega : U(\omega) = \min_{\xi} U(\xi) \} \quad (8.5)$$

e  $\pi_0$  sia la distribuzione uniforme su  $\Omega_0$ , con

$$\Delta = \max_{\omega} U(\omega) - \min_{\omega} U(\omega) \quad (8.6)$$

Vale il seguente

**TEOREMA** (di annealing)[12]

Si assuma che esista un intero  $\tau \geq N$  tale che per ogni  $t$  si abbia

$$S \subseteq \{n_{t+1}, n_{t+2}, \dots, n_{t+\tau}\}$$

e sia  $\{T(t)\}$  ogni sequenza decrescente di temperature per cui vale:

a.  $T(t) \rightarrow 0$  per  $t \rightarrow \infty$

b.  $T(t) \geq N\Delta / \log t \quad \forall t \geq t_0$  con  $t_0 \geq 2$

Allora per ogni configurazione iniziale  $\eta$  in  $\Omega$  e per ogni  $\omega$  in  $\Omega$  vale:

$$\lim_{t \rightarrow \infty} P(X(t) = \omega \mid X(0) = \eta) = \pi_0(\omega)$$

In pratica, questo teorema fornisce uno schema di annealing il quale garantisce che, per ogni configurazione iniziale del processo immagine, il campionatore di Gibbs a temperatura decrescente asintoticamente produce un campione che massimizza la  $\pi$ , cioè di probabilità massima.

## 9. Esempi di probabilità a priori gibbsiane

In questo paragrafo verranno dati alcuni esempi di densità di probabilità a priori gibbsiane per processi immagine modellati come MRF. In accordo alle eqq. (5.2) e (5.3) di paragrafo 5, tale densità di probabilità risulta completamente determinata specificando un sistema di intorno e un insieme di funzioni che attribuiscono un valore di potenziale alle possibili configurazioni assunte da clique di elementi. Questo consente, come già detto, di incorporare nella densità di probabilità congiunta per l'immagine le informazioni a priori possedute sul suo comportamento locale.

Con riferimento all'eq.(5.4), in cui è evidenziata la possibilità di agire separatamente sul processo di intensità e su quello di linea, si descriverà un possibile modello per un processo di intensità costante a tratti, un modello del tutto generale per processi di linea in immagini reali, e un modello che esplicita le interazioni reciproche fra processo di intensità e processo di linea,

legando il processo di linea alla presenza di discontinuità nel processo di intensità.

### Esempio di modello per il processo di intensità

Come già detto, attraverso i termini di potenziale nelle sole variabili  $f$ , è possibile modellare il comportamento locale del campo  $X$  per quanto riguarda il processo di intensità. Per esempio, la classe delle immagini costanti a tratti, cioè le immagini formate dall'unione di regioni a livello di grigio costante, può essere modellata come un MRF sul sistema di intorni del primo ordine di Figg. 1 e 2. Sulla base delle (5.2) e (5.3), per queste immagini il valore di potenziale sulle clique costituite da un solo sito sarà zero, mentre per le clique costituite da due siti,  $C=\{r,s\}$  (coppie di pixel orizzontali e verticali), si avrà:

$$V_C(f) = V_C(f_r, f_s) = \begin{cases} -1 & \text{se } f_r = f_s \\ 1 & \text{altrimenti} \end{cases} \quad (9.1)$$

In accordo a questo modello, le configurazioni di  $F$  "più probabili" (cioè a potenziale più basso) sono quelle in cui due pixel adiacenti hanno lo stesso valore. Quindi il massimo valore per la probabilità congiunta così definita sarà raggiunto per le immagini completamente costanti, indipendentemente dal livello di grigio; tuttavia, l'aver adottato un sistema di intorni molto circoscritto, come quello del primo ordine, fa sì che risultino ad alta probabilità anche immagini costanti solo a tratti.

### Esempio di modello per il processo di linea

Attraverso un'opportuna definizione per la probabilità a priori del processo di linea è possibile, come già visto, vincolare le linee dell'immagine a rispettare, in probabilità, caratteristiche desiderate. Tali caratteristiche saranno, nella maggior parte dei casi, riferite a proprietà locali tipiche quali continuazioni,

terminazioni, incroci, parallelismi e così via, che possono essere descritte dalle configurazioni di gruppi di elementi di linea adiacenti. I vincoli desiderati potranno così essere incorporati in una probabilità a priori gibbsiana attribuendo opportuni valori di potenziale alle suddette configurazioni.

Con riferimento al sistema di intorno di Fig.5, e alle due sole clique riportate in Fig.6, è possibile modellare il processo di linea della maggior parte delle immagini del mondo reale. Le clique a due elementi (orizzontali o verticali) serviranno a modellare parallelismi fra linee, corrispondenti alle configurazioni in cui entrambi gli elementi sono attivi. Volendo attribuire una bassa probabilità al verificarsi dell'evento "linee parallele", si dovrà assegnare un valore di potenziale alto a tali configurazioni. Per le clique a quattro elementi si possono avere le seguenti configurazioni:

1. la configurazione in cui tutti e quattro gli elementi di linea sono disattivi. Ragionevolmente, può essere fatto corrispondere un valore di potenziale minimo, ad esempio zero (massima probabilità) a questa configurazione, in quanto, nella maggior parte delle immagini reali, gli elementi di linea sono pochi e sparsi.
2. le quattro configurazioni in cui un solo elemento di linea è attivo, corrispondenti ad una situazione di fine linea. Un valore di potenziale molto alto indicherà che nelle immagini così modellate le brusche terminazioni di linee risultano poco probabili.



- 3 la configurazione in cui tutti gli elementi di linea sono attivi, corrispondente ad una intersezione fra linee, usualmente a probabilità bassa.
4. le quattro configurazioni in cui sono presenti due elementi di linea attivi ad angolo retto, a probabilità media nelle immagini reali.
5. le quattro configurazioni in cui è presente un elemento a T (tre elementi di linea attivi), ancora a probabilità media.
6. le due configurazioni in cui sono presenti due elementi di linea attivi allineati. In generale a tali configurazioni dovrà essere assegnata una probabilità alta, corrispondendo a continuazioni di linee. Per esempio, la probabilità di queste configurazioni potrà essere complementare a quella assegnata all'evento "fine linea", descritto al punto 2.

#### Esempio di modello di interazione fra pixel e linee

Risulta interessante esaminare adesso un tipico esempio di termine potenziale misto, nelle variabili di linea e di intensità. Come già osservato, un tale termine deve essere determinato in modo che modelli efficacemente l'interazione reciproca fra elementi di linea ed elementi di intensità adiacenti. Le clique miste su cui esso è definito saranno date dall'unione fra clique di intensità e clique di linea. Per esempio si consideri la clique mista costituita da coppie di siti di intensità adiacenti,  $r$  e  $s$ , e dall'elemento di linea posto tra di essi,  $r$ . Si avrà così il seguente tipico potenziale di interazione

$$V_C(f,l) = \alpha (f_r - f_s)^2 (1 - l_r) + \beta l_r \quad (9.2)$$

con  $\alpha$  e  $\beta$  prefissati. Il valore che esso può assumere sulle varie configurazioni per  $f_r$ ,  $f_s$  e  $l_r$  è dato da:

$$V_C(f_r, f_s, l_r=0) = \alpha (f_r - f_s)^2$$

$$V_C(f_r, f_s, l_r=1) = \beta \tag{9.3}$$

Chiaramente, per  $(f_r - f_s)^2 > \beta/\alpha$  il valore di probabilità più alto si ha per  $l_r=1$ , mentre per  $(f_r - f_s)^2 < \beta/\alpha$  il valore di probabilità più alto si avrà per  $l_r=0$ , e sarà tanto più alto quanto più  $(f_r - f_s)^2$  sarà prossimo a zero. In pratica questo significa che le configurazioni più probabili per i tre elementi saranno quella in cui il processo di intensità è costante sulla coppia di pixel in esame, in assenza di processo di linea, e quella in cui il processo di linea è attivo, per alti valori della differenza  $(f_r - f_s)^2$ . Il valore del rapporto  $\beta/\alpha$  ha in questo caso il significato di soglia sul valore  $(f_r - f_s)^2$  oltre il quale si presuppone, da informazioni a priori sulla struttura dell'immagine, che un elemento di linea possa essere ragionevolmente presente. Il termine  $f_r - f_s$  rappresenta la "disparità" fra i valori di pixel adiacenti, ed è in pratica una derivata discreta direzionale.

## 10. Conclusioni

In questo lavoro sono stati analizzati gli aspetti teorici e computazionali dell'approccio probabilistico, o approccio bayesiano, alla soluzione del problema inverso mal-posto della ricostruzione di immagini da dati distorti e affetti da rumore. L'immagine è stata modellata come una coppia  $X=(F,L)$  in cui  $F$  rappresenta il processo di intensità (osservabile) e  $L$  rappresenta il processo di linea (inosservabile), associato ai contorni nell'immagine, particolarmente utile per modellare le discontinuità sempre presenti in una scena. La regolarizzazione del problema della ricostruzione è stata ottenuta riformulandolo come un problema di stima di Massima Probabilità a Posteriori (MAP), in cui la densità di probabilità a priori per la soluzione è data da una distribuzione di Gibbs. Come noto, tale scelta corrisponde a modellare la

soluzione stessa come un campo di Markov (MRF), e consente di incorporare nel problema le conoscenze che si hanno sul comportamento locale della classe di immagini a cui la soluzione appartiene. Particolare interesse è stato rivolto all'esame degli algoritmi proposti in letteratura per superare il problema dei massimi locali, generalmente presenti nei funzionali costo di tipo MAP. Fra questi è stato esaminato in dettaglio il "Campionatore di Gibbs", metodo di rilassamento stocastico basato sull'algoritmo di Metropolis e sull'annealing simulato, che sfrutta le proprietà delle distribuzioni di Gibbs per derivare le probabilità condizionali dei singoli elementi del campo immagine.

## Bibliografia

1. Bertero M., Poggio T.A., Torre V. "Ill-posed Problems in Early Vision", *Proc.IEEE*, Vol.64, N.8, Aug. 1988, pp.869-889.
2. Poggio T.A., Torre V., Koch C. "Computational Vision and Regularization Theory", *Nature*, Vol.317, 26 Sept. 1985, pp.314-319.
3. Poggio T. "Early Vision: From Computational Structure to Algorithms and Parallel Hardware", *Computer Vision, Graphics and Image Processing*, Vol.31, 1985, pp.139-155.
4. Poggio T., Koch C. "Ill-posed Problems in Early Vision: from Computational Theory to Analogue Networks", *Proc. R. Soc. Lond.*, Vol.B 226, 1985, pp.303-323.
5. Maybank S.J. "Algorithm for Analysing Optical Flow Based on the Least-Squares Method", *Image and Vision Computing*, Vol.4, N.1, Feb. 1986, pp.38-42.
6. Trivedi H.P. "Estimation of Stereo and Motion Parameters using a Variational Principle", *Image and Vision Computing*, Vol. 5, N.2, May 1987, pp.181-183.
7. Hildreth E.C. "Computations Underlying the Measurement of Visual Motion", *Artificial Intelligence*, Vol.23, 1984, pp.309-354.
8. Blake A., Zisserman A. *Visual Reconstruction*, The MIT Press, 1987, Cambridge, MA.
9. Blake A., Zisserman A. "Localising discontinuities using weak continuity constraints", *Pattern Recognition Letters*, Vol.6, 1987, pp.51-59.

10. Blake A. "Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction", *IEEE Trans. on Pattern Anal. and Machine Intell.*, Vol.11, N.1, Jan. 1989, pp.2-12.
11. Hunt B.R. "Bayesian Methods in Nonlinear Digital Image Restoration", *IEEE Trans. on Computers*, Vol. C-26, N.3, March 1977, pp.219-229.
12. Geman S., Geman D. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images", *IEEE Trans. on Pattern Anal. and Machine Intell.*, Vol. PAMI-6, N.6, Nov. 1984, pp.721-741.
13. Geman D., Geman S., Graffigne C., Dong P. "Boundary Detection by Constrained Optimization", in stampa, 1988.
14. Marroquin J., Mitter S., Poggio T. "Probabilistic Solution of Ill-posed Problems in Computational Vision", *Journal of the American Statistical Association*, Vol. 82, N. 397, Theory and Methods, March 1987, pp.76-89.
15. Kirkpatrick S., Gelatt C. D., Vecchi M. P. "Optimization by Simulated Annealing", *Science*, Vol. 220, N. 4598, 13 May 1983, pp.671-680.
16. Trussel H. J. "The Relationship Between Image Restoration by the Maximum A Posteriori Method and a Maximum Entropy Method" *IEEE Trans. on ASSP*, Vol. 28, N.1, Feb. 1980, pp.114-117.
17. Besag J. "On the Statistical Analysis of Dirty Pictures", *J. R. Statist. Soc. B*, Vol. 48, N. 3, 1986.
18. Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E. "equations of state calculations by fast computing machines", *J.Chem. Phys.*, Vol.21, 1953, pp.1087-1091.

19. Luenberger D.G. Optimization by vector spaces methods, J. Wiley & Sons, 1969, New York.
20. Luenberger D.G. Linear and nonlinear programming, Addison-Wesley, II Ed., 1984, MA.
21. Scales L.E. Introduction to non-linear optimization, Springer-Verlag, 1985.
22. Tikhonov A.N., Arsenin V.Y. Solutions of ill-posed problems, Winston-Wiley, 1977, Washington.
23. Andrews H.C., Hunt B.R. Digital Image Restoration, Prentice-Hall, 1977, Englewood Cliffs.
24. Dobrushin R.L. "The description of a random field by means of conditional probabilities and conditions of its regularity", *Theory Prob. Appl.*, Vol.13, 1968, pp.197-224.
25. Gull S.F., Daniell G.J. "Image reconstruction from incomplete and noisy data", *Nature*, Vol.272, 20 April 1978, pp.686-690.
26. Wernecke S.J, D'Addario L.R. "Maximum entropy image reconstruction", *IEEE Trans. on Computers*, Vol.26, 1977, pp.351-364.
27. Bertero M. "Linear inverse and ill-posed problems", *Internal Report INFN/TC-88/2*, 1988.
28. Kinderman R., Snell J.L. Markov Random Field and Their Applications, Amer. Math. Soc., 1980, Providence.
29. Leahy R.M., Tonazzini A. "Maximum entropy signal restoration from short data records", *Proc. 8-th IASTED Int. Symp. MECO '86*, Taormina, Italy, 1986.

30. Leahy R.M., Tonazzini A., Wang H. "A dual space optimization technique for maximum entropy signal reconstruction and restoration", *Proc. 20-th Asilomar Conf. on Signals, Systems and Computers*, Monterey, CA, 1986.
31. Bedini L., Tonazzini A. "The use of neural network in maximum entropy image restoration", in stampa, 1989.
32. Marroquin J.L. "Optimal bayesian estimators for image segmentation and surface reconstruction", MIT A.I. Memo N.839, 1985.
33. Torre V., Poggio T.A. "On edge detection", *IEEE Trans. on Pattern Anal. and Machine Intell.*, Vol.8, N.2, 1986, pp.147-163.
34. Norton S.J. "Iterative seismic inversion" *Geophysical Journal*, Vol.94, 1988, pp.457-468.