Mixture models – Soft Tissue Filter

I. Frosio
AIS Lab.
frosio@dsi.unimi.it
http:\\homes.dsi.unimi.it\~frosio\

Overview

- Mixture models: introduzione;
- Segmentazione di radiografia cefalometrica;
- Derivazione di EM per il caso di un mixture model a componenti gaussiane;
- Ottimizzazione vincolata (moltiplicatori di Lagrange);
- Soft Tissue Filter.

Mixture models: introduzione

- La funzione di verosimiglianza permette di stimare i parametri incogniti di una distribuzione (es. media e varianza di una gaussiana);
- Nei casi reali, la densità di probabilità può essere complessa a piacere;
- In particolare, si può considerare una combinazione lineare (mixture) di variabili casuali → **mixture models**

Mixture models: introduzione

- In un mixture model, la d.d.p. complessiva è la combinazione lineare di M d.d.p. di base.

Probabilità del dato x nella J-esima distribuzione

ddp complessiva

Probabilità che il dato x provenga dalla J-esima distribuzione

$$\underbrace{p(\mathbf{x})}_{\text{ddp complessiva}} = \sum_{j=1}^{M} \underbrace{p_\theta(\mathbf{x} \mid j)}_{\text{Probabilità del dato x nella J-esima distribuzione}} \underbrace{P(j)}_{\text{Probabilità che il dato x provenga dalla J-esima distribuzione}}$$

$$\sum_{j=1}^{M} P(j) = 1; \; 0 <= P(j) <= 1; \; \int p_\theta(\mathbf{x} \mid j) = 1$$

Mixture models: introduzione

(incognite)

$$p(\mathbf{x}) = \sum_{j=1}^{M} p_\theta(\mathbf{x} \mid j) P(j)$$

Dati

La d.d.p. complessiva, p(x), viene misurata.

La forma delle d.d.p. di base, $p_q(x|j)$, viene scelta *a priori* (es. gaussiana).

Incognite

I parametri θ di ogni d.d.p. di base, $p_\theta(x|j)$, devono essere stimati.

Le probabilità per ogni d.d.p. di base, P(j), devono essere stimate.

Mixture models: introduzione

- Per la stima del vettore dei parametri θ viene utilizzato…
- … L'approccio alla **massima verosimiglianza**.
- **Mixture model**:
- $p(x) = \Sigma_{j=1..M}\, p_\Theta(x \mid j) \cdot P(j)$
- **Likelihood function**:
- $L = L(\theta) = p(x_1 / \theta) \cdot p(x_2 / \theta) \cdot \ldots\, p(x_D / \theta)$;
- **Negative log likelihood function**:
- $E = E(\theta) = -\log(L) = -\Sigma_{i=1..D} \log [p(x_i / \theta)] =$
  $= -\Sigma_{i=1..D} \log [\Sigma_{j=1..M}\, p_\Theta(x_i \mid j) P(j)]$

Mixture models: esempi

- In un supermercato vi sono due tipi di clienti: single e coppie. La percentuale di single è pari al 33%. La spesa per un single è descritta da una variabile casuale distribuita come una gaussiana, media 70 euro, varianza 30 euro; per una coppia, la media è pari a 100 euro, la varianza 40 euro -> mixture model per descrivere la spesa media (istogramma bimodale);
- Si supponga che esistano altre categorie (famiglie, frequentatore occasionale, ...) -> è sufficiente estendere il modello per includere le nuove categorie...
- Si supponga che la spesa per le famiglie sia ben rappresentata da una variabile casuale con distribuzione chi quadro, media xxx, etc...E' un problema? -> NO! Possiamo includere anche questo caso nel mixture model...

Mixture models: introduzione (riassunto)

- Mixture models: una combinazione lineare di densità di probabilità utilizzata per descrivere la densità di probabilità degli elementi di un vettore di dati misurato.
- Le incognite sono i parametri di ogni ddp del mixture e la probabilità di ogni singola componente del mixture.
- I parametri vengono calcolati massimizzando la funzione di verosimiglianza.

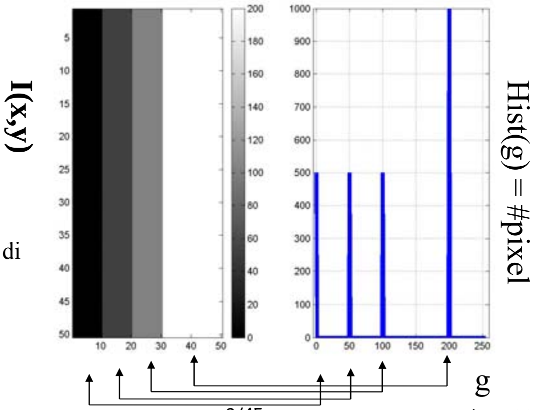
Mixture models: segmentazione radiografia cefalometrica

I(x,y) → immagine NRow x NCol, 8 bit (256 livelli di grigio g);

Hist(·) → istogramma, 256 componenti;

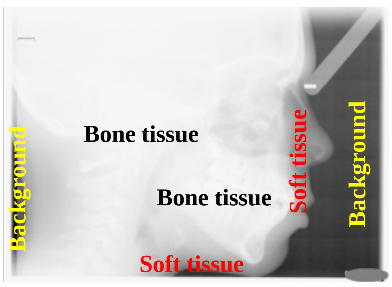
Hist(g) → # pixel t.c. I(x,y)=g;

Hist(g) / (NRow * NCol) = p(g). → **ddp**

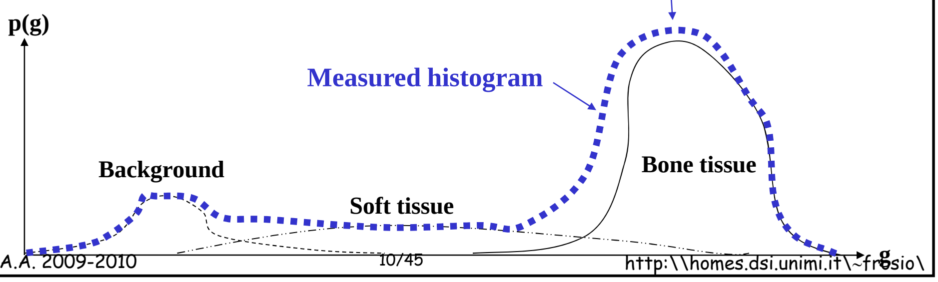


Solo 4 livelli di grigio

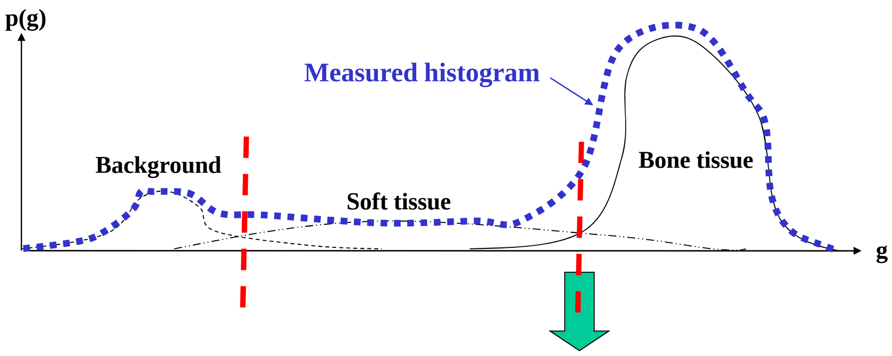
Mixture models: segmentazione radiografia cefalometrica



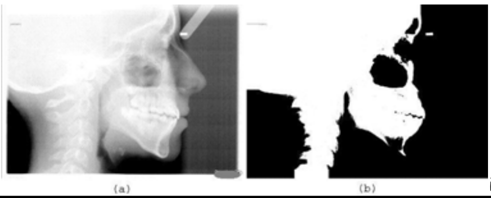
Istogramma tipico immagine radiografica cefalometrica



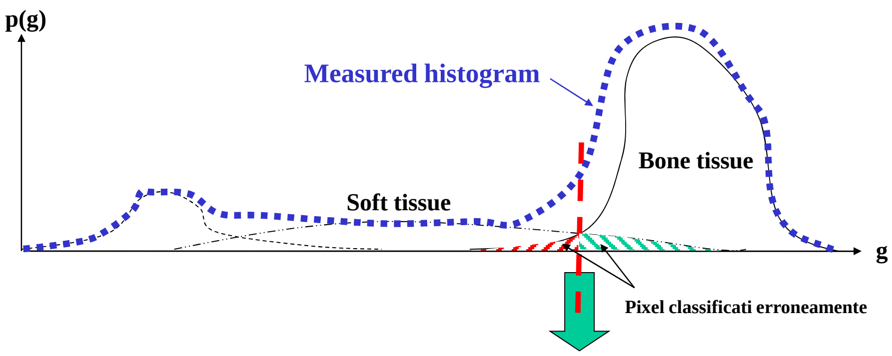
Mixture models: segmentazione radiografia cefalometrica

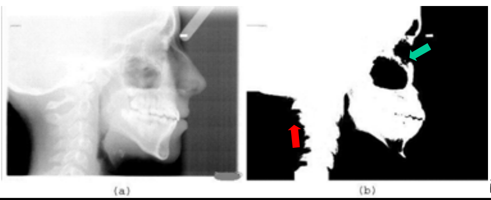


Livelli di grigio diversi identificano tessuti diversi (soglia → segmentazione).



Mixture models: segmentazione radiografia cefalometrica





Mixture models: segmentazione radiografia cefalometrica

- L'istogramma di una radiografia cefalometrica è composto da tre componenti principali;
- Il problema di segmentazione dell'immagine si presta ad essere trattato con un approccio tipo mixture model;
- Ogni componente del mixture model sarà responsabile rispettivamente della generazione dei dati relativi a background, soft tissue e bone tissue;
- Massimizzazione della verosimiglianza: EM (massimizzazione di lower bound).

Mixture models: segmentazione radiografia cefalometrica

- Funzione da minimizzare (incognita θ):

$$E = -\ln(L) = -\ln \prod_{n=1}^{N} p_{\vartheta}(x^n) = -\sum_{n=1}^{N} \ln p_{\vartheta}(x^n)$$

- Per ogni iterazione, i parametri vengono aggiornati (old → new, indice θ omesso):

$$E^{new} - E^{old} = -\sum_{n=1}^{N} \ln p^{new}(x^n) - \left[-\sum_{n=1}^{N} \ln p^{old}(x^n) \right] = -\sum_{n=1}^{N} \ln \left[\frac{p^{new}(x^n)}{p^{old}(x^n)} \right]$$

Mixture models: segmentazione radiografia cefalometrica

$$E^{new} - E^{old} = -\sum_{n=1}^{N} \ln p^{new}(x^n) - \left[-\sum_{n=1}^{N} \ln p^{old}(x^n)\right] = -\sum_{n=1}^{N} \ln\left[\frac{p^{new}(x^n)}{p^{old}(x^n)}\right]$$

Ricordando che:

$$p(x) = \sum_{j=1}^{M} P(j) \cdot p(x \mid j)$$

1

Si ottiene:

$$E^{new} - E^{old} = \sum_{n=1}^{N} -\ln\left[\sum_{j=1}^{M} \frac{P^{new}(j)p^{new}(x^n \mid j)}{p^{old}(x^n)} \cdot \frac{P^{old}(j \mid x^n)}{P^{old}(j \mid x^n)}\right]$$

Mixture models: segmentazione radiografia cefalometrica

- La disuguaglianza di Jensen dice che:

$$dati \quad \lambda_j^2 \quad t.c. \quad \sum_{j=1}^{M} \lambda_j^2 = 1$$

$$\ln\left(\sum_{j=1}^{M} \lambda_j^2 K_j\right) \geq \sum_{j=1}^{M} \lambda_j^2 \ln(K_j) \Rightarrow$$

$$\Rightarrow -\ln\left(\sum_{j=1}^{M} \lambda_j^2 K_j\right) \leq -\sum_{j=1}^{M} \lambda_j^2 \ln(K_j)$$

Mixture models: segmentazione radiografia cefalometrica

$$\sum_{j=1}^{M} \lambda_j^2 = 1 \longleftrightarrow \sum_{j=1}^{M} P^{old}\left(j \mid x^n\right) = 1, \forall n$$

Jensen | Mixture model

E' possibile applicare la disuguaglianza di Jensen ai mixture model, ove i $P^{old}(j \mid x^n)$ giocano il ruolo dei λ_j^2.

Mixture models: segmentazione radiografia cefalometrica

- Applicando Jensen:

$$-\ln\left(\sum_{j=1}^{M} \lambda_j^2 K_j\right) \le -\sum_{j=1}^{M} \lambda_j^2 \ln\left(K_j\right)$$

$$E^{new} - E^{old} = \sum_{n=1}^{N} -\ln\left[\sum_{j=1}^{M} \frac{P^{new}(j) p^{new}\left(x^n \mid j\right)}{p^{old}\left(x^n\right)} \cdot \frac{P^{old}\left(j \mid x^n\right)}{P^{old}\left(j \mid x^n\right)}\right]$$

Mixture models: segmentazione radiografia cefalometrica

$$E^{new} - E^{old} = \sum_{n=1}^{N} -\ln\left[\sum_{j=1}^{M} \frac{P^{new}(j)p^{new}(x^n \mid j)}{p^{old}(x^n)} \cdot \frac{P^{old}(j \mid x^n)}{P^{old}(j \mid x^n)}\right]$$

$$\leq -\sum_{j=1}^{M} P^{old}(j \mid x^n) \cdot \ln\left[\frac{P^{new}(j) \cdot p^{new}(x^n \mid j)}{p^{old}(x^n) \cdot P^{old}(j \mid x^n)}\right]$$

Mixture models: segmentazione radiografia cefalometrica

$$E^{new} - E^{old} \leq -\sum_{n=1}^{N}\sum_{j=1}^{M} P^{old}(j \mid x^n) \ln\left\{\frac{P^{new}(j)p^{new}(x^n \mid j)}{p^{old}(x^n) \cdot P^{old}(j \mid x^n)}\right\}$$

$$E^{new} \leq E^{old} + Q$$

Minimizzando Q si minimizza E!

$P^{old}(x^n)$, $P^{old}(j \mid x^n)$ → costanti…

… dunque $Q = Q(\theta^{new})$!

Mixture models: segmentazione radiografia cefalometrica

Eliminando i termini costanti (old) nella somma (! trasf. Logaritmica!), è sufficiente minimizzare ad ogni iterazione:

$$\widetilde{Q} = -\sum_{n=1}^{N}\sum_{j=1}^{M} P^{old}\left(j \mid x^n\right)\ln\left\{P^{new}\left(j\right)p^{new}\left(x^n \mid j\right)\right\}$$

Tenendo inoltre conto del fatto che:

$$\sum_{j=1}^{M} P^{new}\left(j \mid x^n\right) = 1, \forall n$$

Siamo tornati alla funzione expectation!!! L'ottimizzazione è però, in questo caso, VINCOLATA...

Mixture models: segmentazione radiografia cefalometrica

Si minimizza allora (***metodo dei moltiplicatori di Lagrange***):

$$f = \widetilde{Q} + \psi\underbrace{\left(\sum_{j=1}^{M} P^{new}(j) - 1\right)}_{\textit{Vincolo}}$$

Cioè:

$$\begin{cases} \dfrac{\partial f}{\partial \mu_j^{new}} = 0 \\ \dfrac{\partial f}{\partial \sigma_j^{new}} = 0 \\ \dfrac{\partial f}{\partial P^{new}(j)} = 0 \\ \dfrac{\partial f}{\partial \psi} = 0 \quad \text{(vincolo)} \end{cases} \longrightarrow$$

Da questo sistema possono essere ricavate le equazioni per l'aggiornamento dei parametri del mixture model ad ogni iterazione.

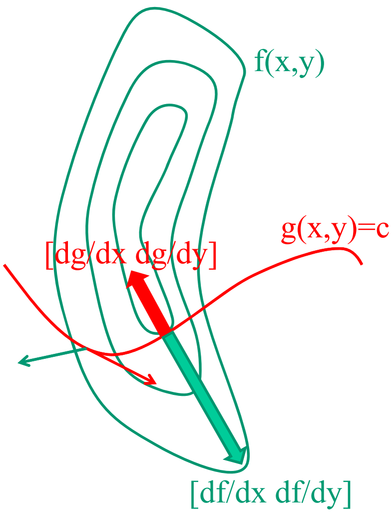
Ottimizzazione vincolata (moltiplicatori di Lagrange)

- Vogliamo trovare (x,y) t.c. f(x,y) è minimia nell'insieme g(x,y) = c.
- f(x,y) -> funzione da ottimizzare.
- g(x,y) – c = 0 -> vincolo.

- Nel caso dell'ottimizzazione non vincolata -> (df/dx = 0) & (df/dy = 0).

Ottimizzazione vincolata (moltiplicatori di Lagrange)

- Nel caso di ottimizzazione vincolata → la g(x,y) - c =0 identifica una linea curva nello spazio (x,y).
- Il vettore [dg/dx dg/dy] identifica la normale a g(x,y)=c.
- Il punto massimo/minimo t.x. g(x,y) = c deve essere t.c. [dg/dx dg/dy] è parallelo al gradiente di f [quindi, muovendosi lungo g(x,y)=c, si osserva localmente una variazione nulla di f].



Ottimizzazione vincolata (moltiplicatori di Lagrange)

- Coma facciamo ad imporre che [dg/dx dg/dy] sia parallelo a [df/dx df/dy]?
- Imponiamo che grad(f) possa essere ottenuto semplicemente scalando grad(g) per un fattore λ (incognito - moltiplicatore di Lagrange), ovvero:

  $\text{grad}(f) + \lambda \cdot \text{grad}(g) = 0 \qquad (1)$
- L'equazione (1) si ottiene **minimizzando** (ottimizzazione non vincolata) **la funzione** $f+\lambda\cdot(g-c)$ **(detta funzione Lagrangiana)**.

Ottimizzazione vincolata (moltiplicatori di Lagrange)

- $(x^*, y^*, \lambda^*) = \text{argmax}_{x,y,\lambda}\ f(x,y)+\lambda\cdot(g(x,y)-c)$ =>

$$\begin{cases} df(x^*,y^*)/dx+\lambda\cdot dg(x,y)/dx = 0 \\ df(x^*,y^*)/dy+\lambda\cdot dg(x,y)/dy = 0 \\ g(x,y)-c = 0 \end{cases}$$

Derivando rispetto a x,y ottieniamo il parallelismo tra i due gradienti

Derivando rispetto a λ ottieniamo il vincolo

Mixture models: segmentazione radiografia cefalometrica

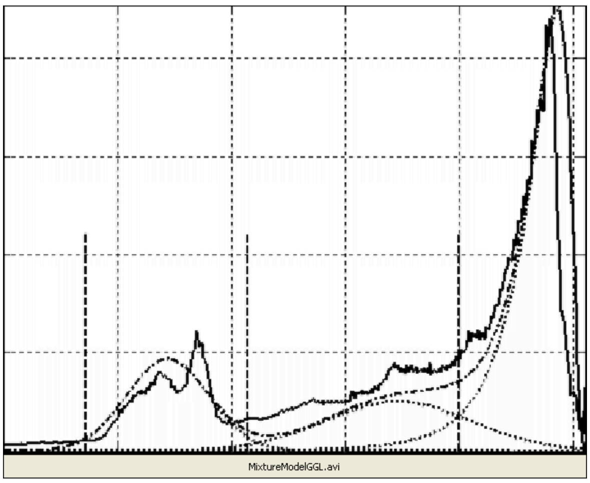
- Aggiornamento P(j):

$$P^{new}(j) = \frac{1}{N}\sum_{n=1}^{N} P^{old}\left(j \mid x^n\right)$$

- Nel caso di ddp gaussiane:

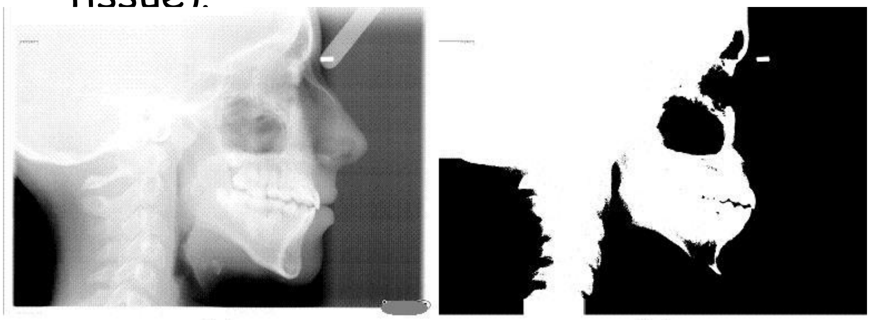
$$\mu_j^{new} = \frac{\sum_{n=1}^{N} P^{old}\left(j \mid x^n\right)x^n}{\sum_{n=1}^{N} P^{old}\left(j \mid x^n\right)} \qquad \left(\sigma_j^{new}\right)^2 = \frac{\sum_{n=1}^{N} P^{old}\left(j \mid x^n\right)\left(x^n - \mu_j\right)^2}{\sum_{n=1}^{N} P^{old}\left(j \mid x^n\right)}$$

Mixture models: segmentazione radiografia cefalometrica



Risultato (bone tissue)

- Clusterizzazione immagine in 3 zone (background, soft tissue, bone tissue).



(a) (b)

Bibliografia

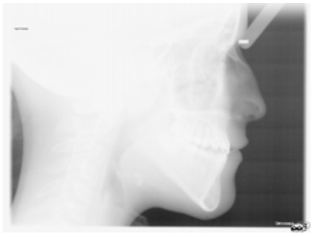
- Cristopher M. Bishop, Pattern Recognition and Machine Learning, Capitolo 2.3.9 (mixture di gaussiane), Capitolo 9.1, 9.2, 9.3 (K-means, mixture models, EM).

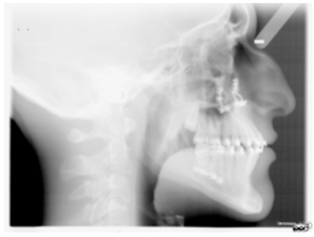
- I. Frosio, G. Ferrigno, N. A. Borghese, "Enhancing Digital Cephalic Radiography with Mixture Model and Local Gamma Correction," IEEE Transaction on Medical Imaging, Vol. 25, No. 1, Jan. 2006, pp. 113-121 (mixture models e radiografia cefalometrica).

- *Poul Erik Frandsen, Kristian Jonasson, Hans Bruun Nielsen, Ole Tingleff, Unconstrained Optimization, disponibile in rete, algoritmi di minimizzazione (approfondimento).*

- *Cristopher M. Bishop, Pattern Recognition and Machine Learning, Capitolo 9.4, EM come minimizzazione di un lower bound (approfondimento).*

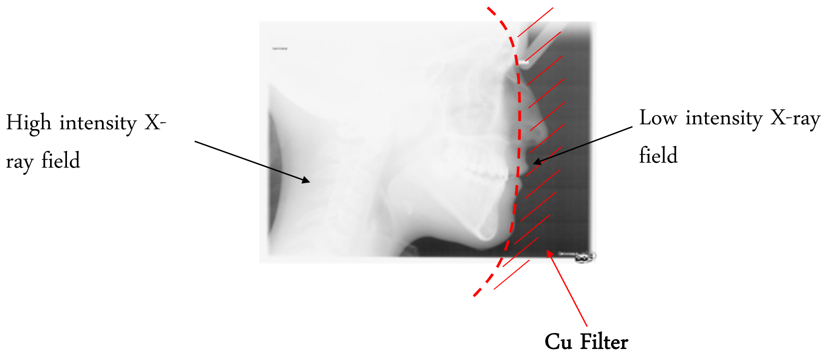
Exposure problems





- From film to phosphor film to CCD
- Underexposed radiographies: bone cannot be distinguished from soft tissue
- Overexposed radiographies: soft tissue tends to mix with background

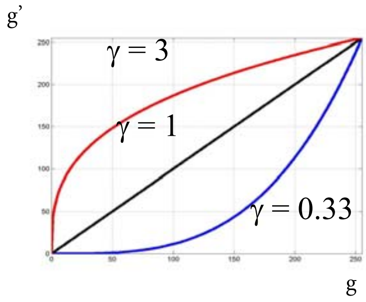
Typical HW solution



SW solution: gamma correction

- Gray level correction formula;
- Used for global exposure correction;
- For 8 bit image, gray level g is corrected to g' as:

  $g' = 255 \cdot [(g / 255)^{1/\gamma}]$



$\gamma > 1 \rightarrow$ Stretching of the low levels, compression of the high levels.

$\gamma < 1 \rightarrow$ Compression of the low levels, stretching of the high levels.

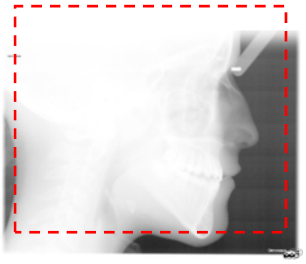
Global γ correction

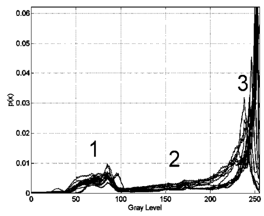
γ = 3

γ = 0.25

- Nowadays clinical practice: global γ correction
- γ = 0.25 (underexposed), soft tissue darkens!
- γ = 3 (overexposed), bone tissue mixes with soft tissue!

Typical histogram





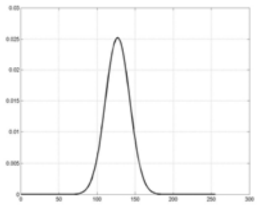
- Three characteristic gray zones: background (1), soft tissue (2), bone tissue (3)
- 5% boundary eliminated (white margins, logo)
- Gray level zero eliminated (saturated pixels)

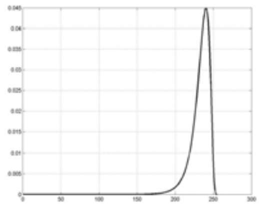
Mixture model

- A powerful technique to estimate complex probability densities distributions by using a restricted number of parameters
- Linear combination of M probability densities:

$$p_{MM}(x) = \sum_{j=1}^{M} P(j) \cdot p(x \mid j)$$

Two Gaussian, one Lognormal





- Mixture of three components (M = 3)
- Two Gaussians: background, soft tissue (symmetric peaks)
- One Inverted Lognormal: bone tissue (asymmetric peak)

Parameters estimation

- Parameters: P(j), μ_j, σ_j, j=1…3
- Negative log likelihood

$$E = -\ln L = -\sum_{n=1}^{N} \ln p_{MM}(x^n) = -\sum_{n=1}^{N} \ln\left\{p(x^n \mid j)P(j)\right\}$$

- E is minimized through the EM algorithm

Updating equations

- Mixing parameters $P(j)^{new} = \frac{1}{N} \sum_{g=0}^{N_{GL}-1} P^{old}(j \mid g) \cdot H(g)$

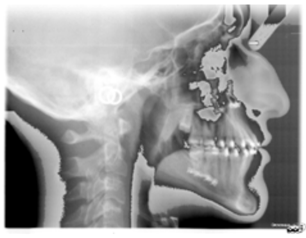
- Gaussians $\mu_j^{new} = \frac{\sum_{g=0}^{N_{GL}-1} P^{old}(j \mid g) \cdot g \cdot H(g)}{\sum_{g=0}^{N_{GL}-1} P^{old}(j \mid g) \cdot H(g)}$ $\left(\sigma_j^{new}\right)^2 = \frac{\sum_{g=0}^{N_{GL}-1} P^{old}(j \mid g) \cdot \left(g - \mu_j^{new}\right)^2 \cdot H(g)}{\sum_{g=0}^{N_{GL}-1} P^{old}(j \mid g) \cdot H(g)}$

- Lognormal $\mu_j^{new} = \frac{\sum_{g=0}^{N_{GL}-1} P^{old}(j \mid g) \cdot \ln(N_{GL} - g) \cdot H(g)}{\sum_{g=0}^{N_{GL}-1} P^{old}(j \mid g) \cdot H(g)}$ $\left(\sigma_j^{new}\right)^2 = \frac{\sum_{g=0}^{N_{GL}-1} P^{old}(j \mid g) \cdot \left[\ln(N_{GL} - g) - \mu_j^{new}\right]^2 \cdot H(g)}{\sum_{g=0}^{N_{GL}-1} P^{old}(j \mid g) \cdot H(g)}$

Clustering

- Thresholds *Th(j,j+1)* minimize
$$\int_0^{Th(j,j+1)} P(j+1) \cdot p(x \mid j+1) dx + \int_{Th(j,j+1)}^{N_{GL}-1} P(j) \cdot p(x \mid j) dx$$
- Three classes: Background, soft tissue, bone tissue

Local γ Correction



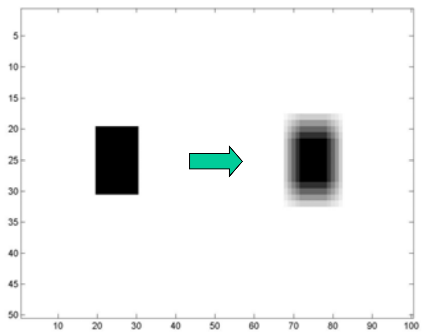
- γ = 1 background
- γ = 0.25 bone tissue
- γ = 1.5 soft tissue
- Artifacts, patient profile not clearly visible!

γ smoothing: low pass filtering

- Low pass filtering of γ …

Filter (5x5):

$$\begin{pmatrix} 1/25 & 1/25 & 1/25 & 1/25 & 1/25 \\ 1/25 & \dots & & & \\ \dots & & & & \\ \dots & & & & 1/25 \end{pmatrix}$$



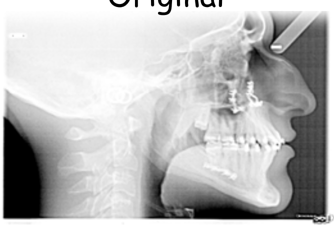
γ smoothing

- γ map has to be smoothed
- Down sampling, moving average 3x3 filtering, up sampling using bilinear interpolation (or efficient moving average filter in space domain)
- Two classes: Background & soft tissue, bone tissue

Results

Original

Gamma correction



Unsharp masking

Soft tissue filter

Summary Soft Tissue Filter

- Mixture model for clustering;
- Gamma correction;
- Low pass filtering of gamma map.