A Marginalized MAP Approach and EM Optimization for Pair-Wise Registration

William Wells

Associate Professor of Radiology
Surgical Planning Laboratory
Harvard Medical School and Brigham and Women's Hospital

Affiliated Faculty: Harvard – MIT Division of Health Sciences and Technology

Research Scientist – MIT CSAIL

A Marginalized MAP Approach and EM Optimization for Pair-Wise Registration

IPMI 2007

Lilla Zollei
Mark Jenkinson
Samson Timoner
William Wells

Outline

- Basic MAP approach to image registration
- Parametric models on image and intensity pairs
- MAP with known model parameters *
- MAP with unknown model parameters
  - Prior probabilities on model parameters
  - Joint MAP *
  - Marginalized MAP
    - Weak prior *
    - Informative prior *
    - Strong prior *
  - EM Algorithm to obtain estimates
    - Simple iteration *
- Experimental Results

* Connections to prior work

Basic MAP Registration

$$\hat{T} = \arg\max_{T} \log p(T|u, v)$$

$$\hat{T} = \arg\max_{T} \log \left[p(u, v|T)\ p(T) \right]$$

- *u*, *v*: images
- *T*: transform on image
- Maximum A-posteriori Probability

Probability on Image Pairs

- *Kinematic* Assumption

$$p(u, v|T) = p(u(x), v(T(x)))$$

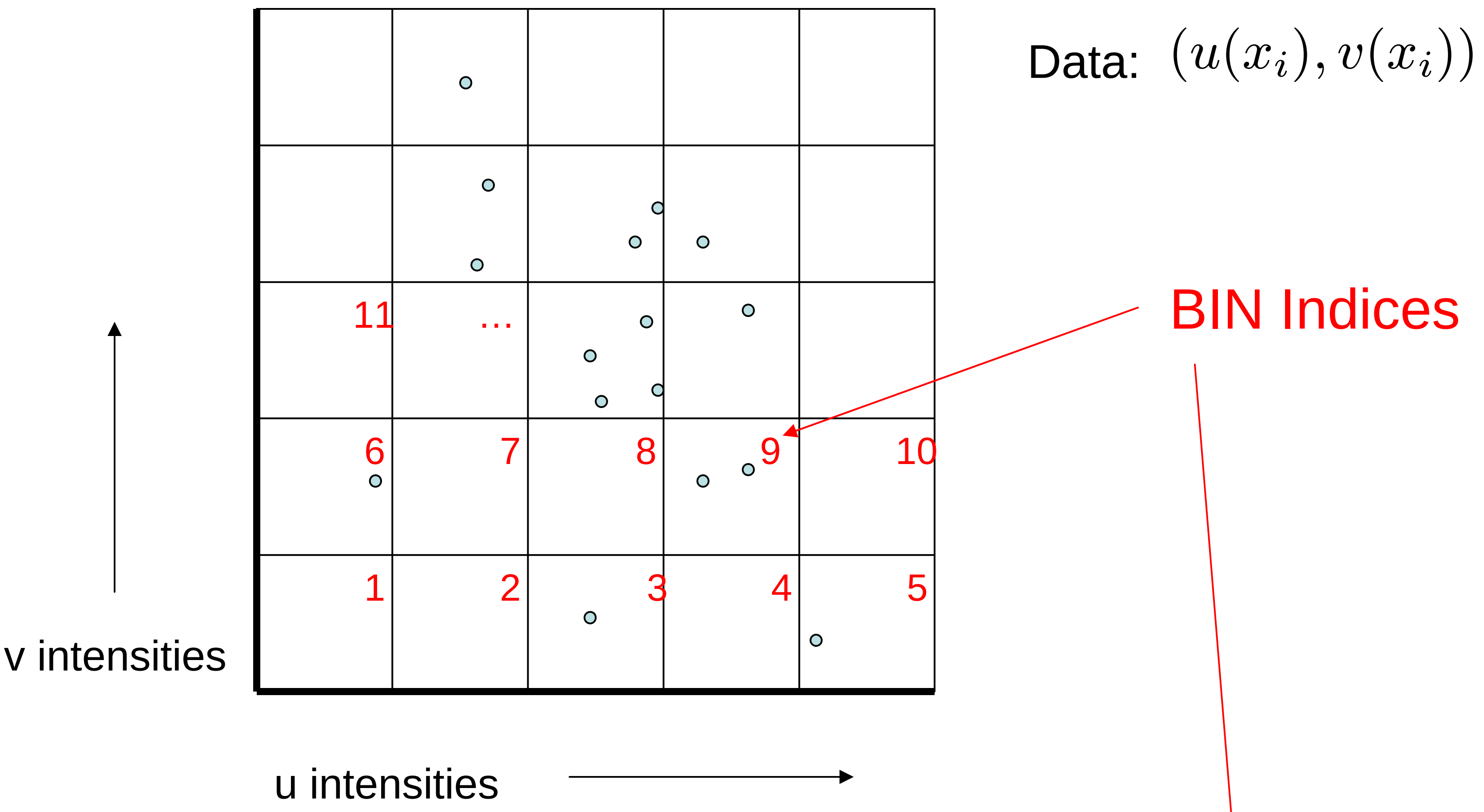
- Independently and Identically Distributed (IID) in space

$$p(u, v|T) = \prod_{x_i} p(u(x_i), v(x_i)|T)$$

- x_i : voxel

Joint Image Histogram

Data: $(u(x_i), v(x_i))$



Bin Counts, e.g., $n_9 = 2$

Probability on Intensity Pairs

$$\begin{aligned} p(u_i, v_i|\Theta) &= \mathrm{CAT}(\mathcal{B}(u_i, v_i); \Theta) \\ &= \theta_{\mathcal{B}(u_i, v_i)} \end{aligned}$$

$$0 \leq \theta_j \leq 1 \qquad \sum_j \theta_j = 1$$

- $\mathcal{B}(\cdot, \cdot)$: Bin index
- CATegorical Distribution

Some Discrete Distributions

	Single trial	Multiple trials
Binary data	*Bernoulli*	*Binomial*
Number of data values > 2	•*Multinomial (in one trial)* •*Categorical* •*PMF* •*Probability Distribution*	*Multinomial*

Probability on Image Pairs…

$$
\begin{aligned}
p(u, v|T, \Theta) & \\
&= \prod_i \mathrm{CAT}(\mathcal{B}(u_i, v_i); \Theta) \\
&= \mathrm{MULT}(\{\mathcal{B}(u_{x_1}, v_{y_1}) \cdots \mathcal{B}(u_{x_N}, v_{y_N})\}; \Theta) \\
&= \prod_j \theta_j^{n_j(T)}
\end{aligned}
$$

- *$n_j(T)$:* Number of voxel pairs that map to bin j
- Multinomial on observations (not counts)

MAP Registration: Known Model

$$\begin{aligned}\hat{T} &= \arg\max_T \log p(T|u,v,\Theta) \\ &= \log\left[p(u,v|T,\Theta)\ p(T)\right] \\ &= \arg\max_T \left[\sum_{j=1}^{g} n(T)_j \log(\theta_j) + \log P(T)\right]\end{aligned}$$

- Without *P(T)* : Maximum Likelihood *

* M Leventon, W Grimson, W Wells. Multi-Modal Volume Registration Using Joint Intensity Distributions. MICCAI 98

MAP Registration: Known Model…

$$\hat{T} \quad = \quad \arg\max_{T} \left[\sum_{j=1}^{g} n(T)_j \log(\theta_j) + \log P(T) \right]$$

- Training
  - Estimate Θ from registered images
    - ML: normalized histogram
- Registration
  - Simple Objective Function
    - Linear in counts
- Capture / Bias issue
  - Model may be inaccurate for new images

Θ Unknown

- Prior: p(Θ|w)
- Joint prior is independent
  - p(T,Θ) = p(T) p(Θ|w)

- Joint Posterior:

$$p(T, \Theta | u, v, w) \propto p(u, v | T, \Theta) p(T) p(\Theta | w)$$

Joint MAP*

$$\widehat{T\Theta} = \arg\max_{T\Theta} p(T, \Theta | u, v, w)$$

- Estimate T, Θ jointly
- Θ is a *nuisance* parameter
  - discard estimate of Θ

$$\widehat{T} = \arg\max_{T} \left[\max_{\Theta} p(T, \Theta | u, v, w) \right]$$

- Θ is *maximized out*

* L . Zöllei. A Unified Information Theoretic Framework for Pair- and Group-wise Registration of Medical Images. Ph.D. thesis, MIT

Joint MAP*

- Θ is *maximized out*

$$\widehat{T} = \arg\max_{T} \left[\max_{\Theta} p(T, \Theta | u, v, w) \right]$$

* L . Zöllei. A Unified Information Theoretic Framework for Pair- and Group-wise Registration of Medical Images.
Ph.D. thesis, MIT

- Joint Maximum Likelihood
  - Connections to entropy and Mutual Information

A. Roche, G. Malandain, and N Ayache. Unifying maximum likelihood approaches in medical image registration. *International Journal of Imaging Systems and Technology*, 11(7180):71–80, 2000

Marginalize Nuisance Parameter

$$\widehat{T} = \arg\max_{T} \left[\max_{\Theta} p(T, \Theta | u, v, w) \right]$$

- Alternative: Average or *Marginalize*:

$$\begin{aligned} \widehat{T} &= \arg\max_{T} \left[\int p(T, \Theta | u, v, w) d\Theta \right] \\ &= \arg\max_{T} p(T | u, v, w) \\ &= \arg\max_{T} \left[\int p(u, v | T, \Theta) p(T) p(\Theta | w) d\Theta \right] \end{aligned}$$

Dirichlet Prior on Θ

$$p(\Theta|w) = \mathrm{DIR}(\Theta; w)$$

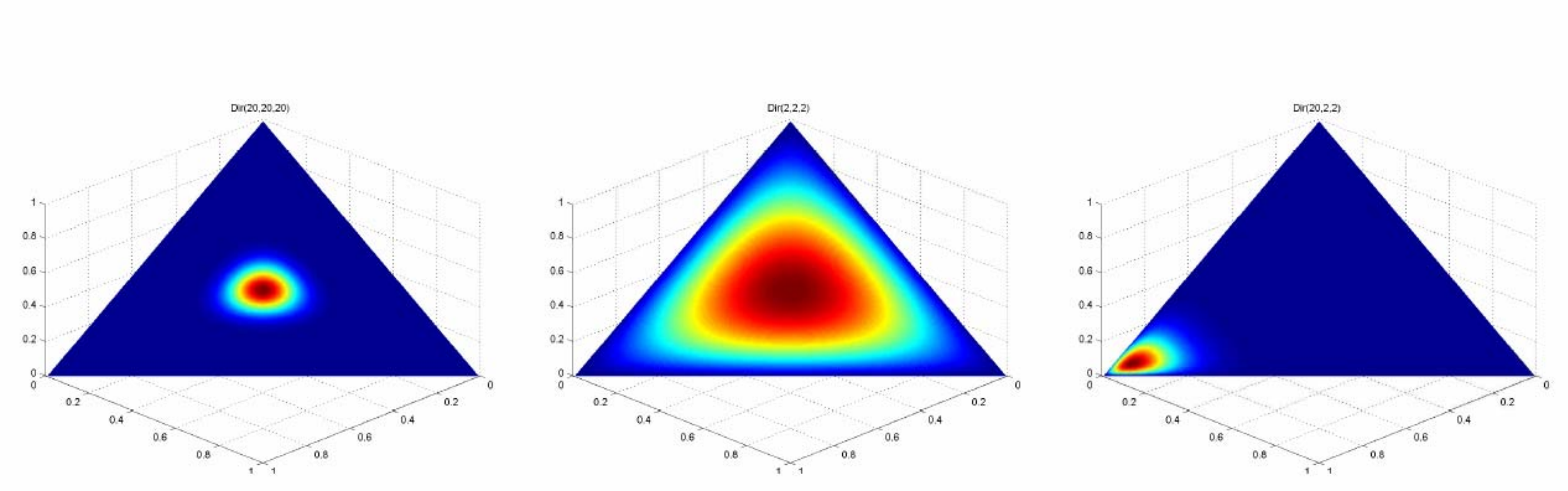
$$= \frac{1}{Z(w)} \prod_{i=1}^{g} \theta_i^{(w_i - 1)} = \Gamma(w_0) \prod_j \frac{\theta_j^{(w_j - 1)}}{\Gamma(w_j)}$$

$$w_0 = \sum_i w_i$$

- Conjugate prior for Multinomial
- Multi-category generalization of Beta
- Parameterized by pseudo-data counts

Examples: order 3 Dirichlet

$$0 \leq \theta_i \leq 1 \qquad \theta_1 + \theta_2 + \theta_3 = 1$$



Bayesian statistics: a concise introduction. Kevin P. Murphy
(tech note on web, use google)

Dirichlet distribution…

- Conjugate property:

$$
\begin{aligned}
p(\Theta|x) &\propto \mathrm{MULT}(x;\Theta)\cdot\mathrm{DIR}(\Theta;w) \\
&\propto \mathrm{DIR}(\Theta;n(x)+w)
\end{aligned}
$$

$$\Theta \sim \mathrm{DIR}(\mathrm{w}) \rightarrow \mathrm{MODE}(\Theta)_j = \frac{w_j - 1}{w_0 - 1}$$

Discrete Distributions

	Single trial	Multiple trials	Conjugate Prior
Binary data	*Bernoulli*	*Binomial*	*Beta*
Number of data values > 2	•*Multinomial (in one trial)* •*Categorical* •*PMF* •*Probability Distribution*	*Multinomial*	*Dirichlelt*

Marginalized Posterior

$$p(T|u,v,w)$$

$$\propto \int p(u,v|T,\Theta)p(T)p(\Theta|w)d\Theta$$

$$\propto P(T)\int \left(\prod_i \theta_i^{n_i(T)}\right)\Gamma(w_0)\left(\prod_j \frac{\theta_j^{(w_j-1)}}{\Gamma(w_j)}\right)d\theta$$

$$\propto P(T)\frac{\Gamma(w_0)}{\Gamma(N+w_0)}\left(\prod_j \frac{\Gamma(n_j(T)+w_j)}{\Gamma(w_j)}\right)$$

Marginalized MAP

$$\begin{aligned}
\hat{T} &= \arg\max_{T} \log p(T|u, v, w) \\
&= \arg\max_{T} \left[\log P(T) + \sum_{j=1}^{g} \log \Gamma\left(n_j(T) + w_j\right) \right]
\end{aligned}$$

3 Cases on strength of prior

- Weak prior: Laplace Prior
- Informative prior
- Strong prior: dominates data

Laplace Prior: $w_j = 1$

$$\hat{T} = \arg\max_T \left[\log P(T) + \sum_{j=1}^{g} \log \Gamma\left(n_j(T) + 1\right) \right]$$

$$\log(\Gamma(x+1)) \approx x\ log(x) - x$$

$$\hat{T} \approx \arg\max_T \left[\log P(T) + \sum_{j=1}^{g} n_j(T) \log n_j(T) \right]$$

$$\hat{T} \approx \arg\min_T \left[N \cdot \mathrm{H}\left[\mathrm{CAT}\left(\frac{n(T)}{N} \right) \right] - \log p(T) \right]$$

Minimum Entropy Registration

- VBC 94 conference antecedent
  - D. **Hill**, **Studholme**, C., and **Hawkes**, D.. Voxel Similarity Measures for Automated Image Registration. VBC 1994
- Entropy for registration
  - **Collignon**, A., **Vandermuelen**, D., **Suetens**, P., and **Marchal**, G.. 3d multi-modality medical image registration using feature space clustering. CVRMED 1995.
- Mutual Information
  - **Viola**, P. and **Wells**, W.. Alignment by maximization of mutual information. In Proceedings of the 5th International Conference of Computer Vision, 1995.
  - **Viola**, P. Alignment by maximization of Mutual Information. MIT PhD Thesis, 1995.

Maximum Mutual Information Registration

$$I(U,V) = H(U) + H(V) - H(UV)$$

- **Wells** WM, **Viola** P, **Atsumi** H, **Nakajima** S, **Kikinis** R. Multi-Modal Volume Registration by Maximization of Mutual Information. Medical Image Analysis, 1(1):35-51, 1996.
- **Viola** P, **Wells** WM. Alignment by maximization of mutual information. International Journal of Computer Vision. 1997;24:137-154.
- F. **Maes**, A. **Collignon**, D. **Vandermeulen**, G. **Marchal**, P. **Suetens**, Multimodality image registration by maximization of mutual information, IEEE transactions on Medical Imaging, vol. 16, no. 2, pp. 187-198, April 1997
- **West** JB, **Fitzpatrick** JM, et al.. "Comparison and evaluation of retrospective intermodality image registration techniques. JCAT 1997.
- Josien P. W. **Pluim**, J. B. Antoine **Maintz**, Max A. **Viergever**: Mutual Information Based Registration of Medical Images: A Survey. IEEE Trans. Med. Imaging 22(8): 986-1004 (2003)

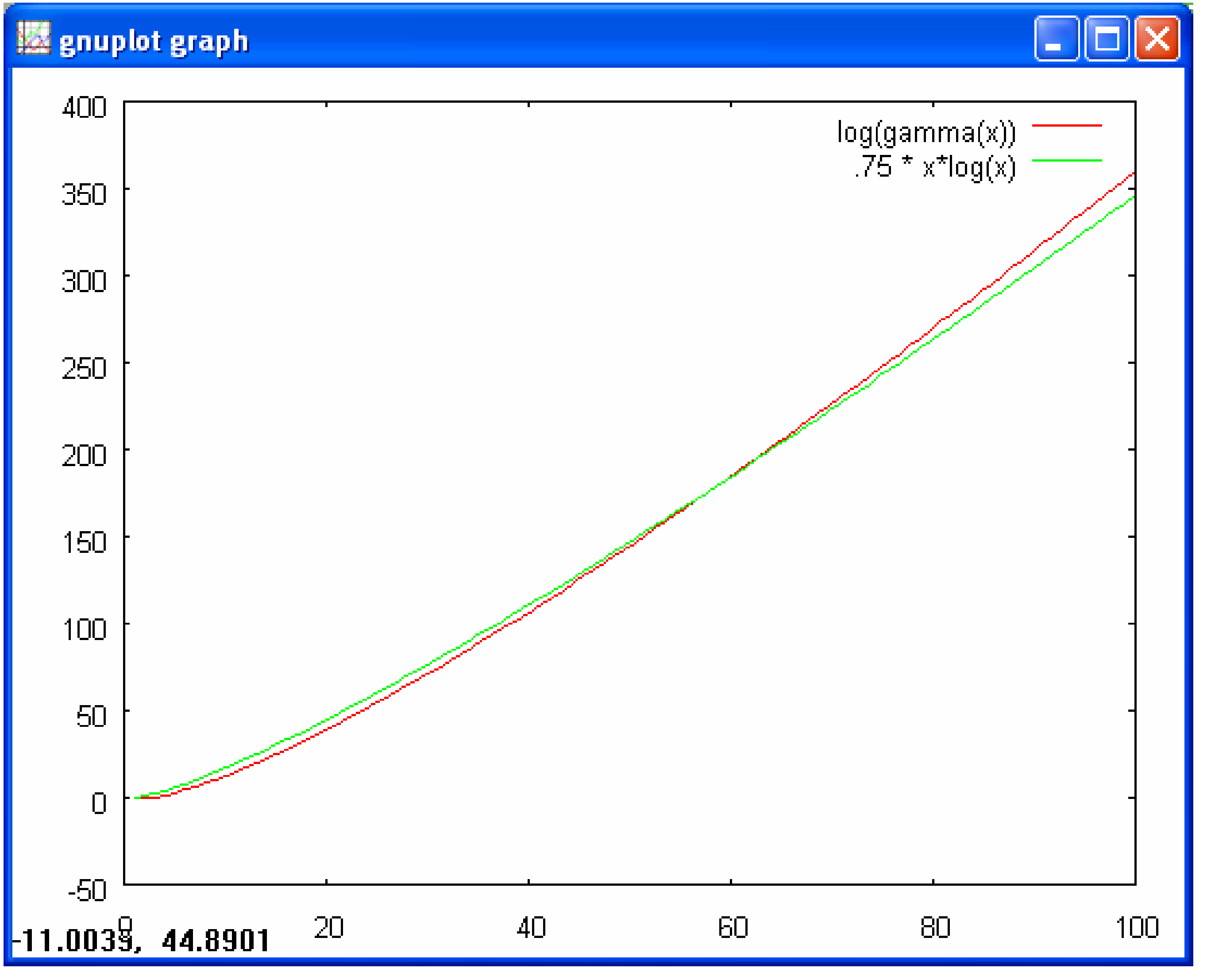
Informative Prior

- Return to Marginal MAP estimator:

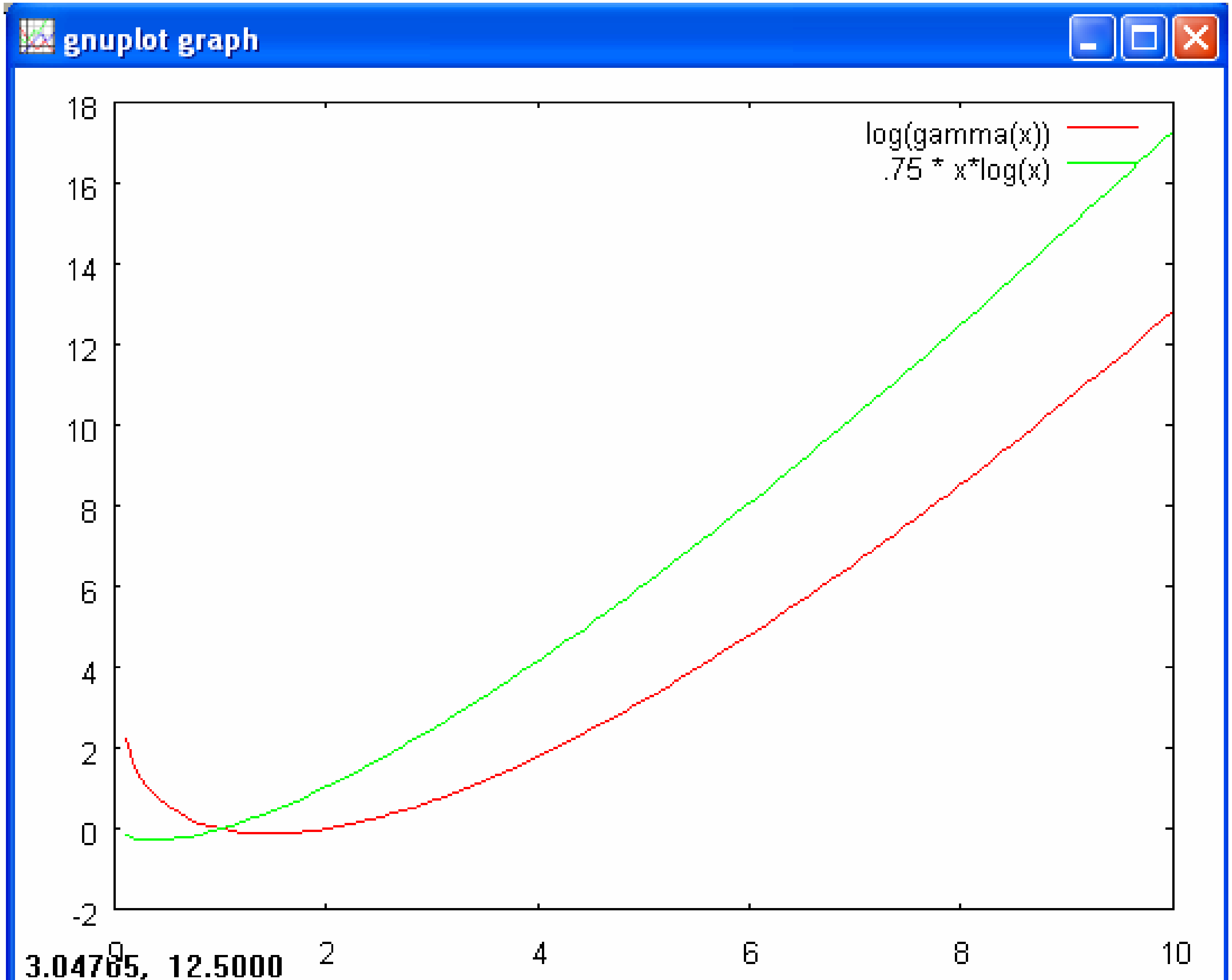
$$\hat{T} \quad = \quad \arg\max_{T} \left[\log P(T) + \sum_{j=1}^{g} \log \Gamma\left(n_j(T) + w_j\right) \right]$$

- Re-examine log Γ(x) …

log (Γ(x)) and *x log (x)*



log (Γ(x)) and *x log (x)*



Informative Prior

$$\hat{T} = \arg\max_T \left[\log P(T) + \sum_{j=1}^{g} \log \Gamma \left(n_j(T) + w_j \right) \right]$$

$$\approx \arg\max_T \left[\log P(T) + c \cdot \sum_{j=1}^{g} (n_j(T) + w_j) \log(n_j(T) + w_j) \right]$$

$$\approx \arg\min_T \left[N \cdot c \cdot \mathrm{H} \left[\mathrm{CAT} \left(\frac{n(T) + w}{N + w_0} \right) \right] - \log p(T) \right]$$

- Minimize entropy of *pooled data* *

* Mert Sabuncu. Entropy-based Methods for Image Registration. PhD Thesis, Princeton 2006.

Strong Prior

$$\hat{T} = \arg\max_{T} \left[\log P(T) + \sum_{j=1}^{g} \log \Gamma \left(n_j(T) + w_j \right) \right]$$

$$\hat{T} \approx \arg\max_{T} \left[\log P(T) + \sum_{j} n_j(T) \log(w_j) \right]$$

- MAP with known model: Θ = w

Marginalized MAP registration

- 3 Cases on strength of prior
  - Weak prior: Laplace Prior
    - Minimize entropy
  - Informative prior
    - Minimize entropy of pooled data
  - Strong prior: dominates data
    - MAP using known fixed model

Outline

- Basic MAP approach to image registration
- Parametric models on image and intensity pairs
- MAP with known model parameters *
- MAP with unknown model parameters
  - Prior probabilities on model parameters
  - Joint MAP *
  - Marginalized MAP
    - Weak prior *
    - Informative prior *
    - Strong prior *
  - EM Algorithm to obtain estimates
    - Simple iteration *
- Experimental Results

EM Algorithm: ML version

$$\hat{\phi} = \arg\max_{\phi} \log p(y|\phi)$$

- Missing data: ψ

$$\hat{\phi}^{n+1} = \arg\max_{\phi} \mathrm{E}_{\psi|y\hat{\phi}^n}[\log p(y\psi|\phi)]$$

$$p(\psi|y\phi) = \frac{p(y\psi|\phi)}{\sum_{\psi} p(y\psi|\phi)}$$

A. P. Dempster, N. M. Laird, and D. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38,1977.

EM Algorithm: ML version

- Missing parameter, independent prior

$$p(y\phi|\psi) = p(y|\psi\phi)p(\psi)$$

$$p(y|\phi) = \int p(y\psi|\phi)d\psi = \int p(y|\psi\phi)p(\psi)d\psi$$

$$\hat{\phi}^{n+1} \quad = \quad \arg\max_{\phi} \mathrm{E}_{\psi|y\hat{\phi}^n}[\log p(y|\psi\phi)]$$

$$\phi^n \to \arg\max_{\phi} \log p(y|\phi)$$ (locally)

EM Algorithm: MAP version

$$\hat{\phi} \quad = \quad \arg\max_{\phi} \log p(\phi|y)$$

$$\hat{\phi}^{n+1} = \arg\max_{\phi} \mathrm{E}_{\psi|y\hat{\phi}^n}[\log p(y|\psi\phi) \quad + \log p(\phi)]$$

EM Estimator of *T* | *u v w*

$$\begin{aligned}\hat{T}_{\text{next}} &= \arg\max_{T} \mathrm{E}_{\Theta|u,v,\hat{T}_{\text{old}}} \left[\log P(u, v, |T, \Theta) + \log P(T) + \log P(\Theta)\right] \\ &\vdots \\ &\approx \arg\max_{T} \left[\sum_{j=1}^{g} n(T)_j \log\left(n(\hat{T}_{\text{old}})_j + w_j - .5\right) + \log P(T)\right]\end{aligned}$$

- Iterate:
  - (Re) estimate model from current configuration
    - Histogram joint intensities
  - Do MAP registration with fixed model
    - Simple objective function: linear in counts

EM Estimator of *T* | *u v w*

$$\hat{T}_{\text{next}} \approx \arg\max_T \left[\sum_{j=1}^{g} n(T)_j \log \left(n(\hat{T}_{\text{old}})_j + w_j - .5 \right) + \log P(T) \right]$$

- Samson's Iteration:

$$\hat{T}_{\text{next}} \approx \arg\max_T \left[\sum_{j=1}^{g} n(T)_j \log \left(n(\hat{T}_{\text{old}})_j + \epsilon \right) + \log P(T) \right]$$

Experiment *

- Samson Timoner PhD Thesis
- Sequential Intra-Operative MRI
- Iterated MAP
- Linear Elastic Deformation Energy: E(T)
  - $p(T) \propto \exp(-E(T))$
- Iterated relaxation of deformation energy
  - Equivalent to Viscous fluid **

* S Timoner. Compact Representations for Fast Nonrigid Registration of Medical Images. PhD Thesis, MIT 2003

** X. Papademetris, E. T. Onat, A. J. Sinusas, D. P. Dione, R. T. Constable, and J. S. Duncan. The active elastic model. In *Proceedings of IPMI*, volume 0558 of *LNCS*, pages 36–49. Springer, 2001.

Samson’s Iteration

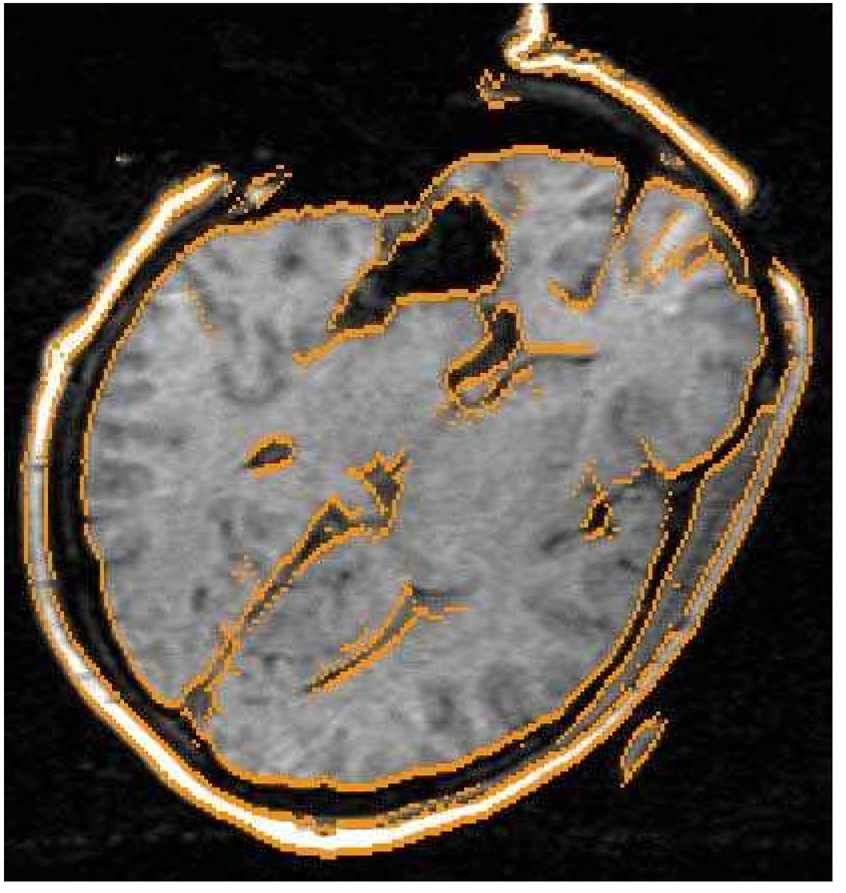
- Iterate:
  - (Re) estimate model from current configuration
    - Histogram joint intensities
  - Do MAP registration with fixed model
    - Simple objective function: linear in counts
  - Relax energy of current deformation

Signa SP (GE Medical Systems)

R. Pergolizzi

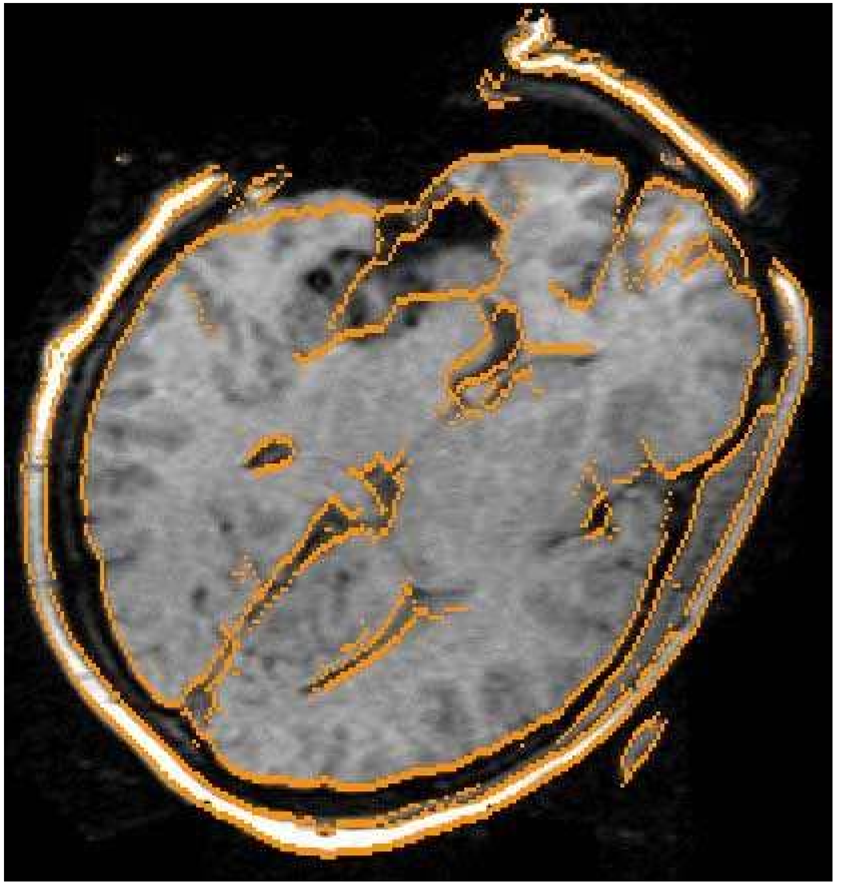
Provided by D. Gering

Intra-Operative Image



Provided by Samson Timoner

Second Intra-Operative Image



Provided by Samson Timoner

Second Image Warped to First Image

Provided by Samson Timoner

Summary

- Described Basic MAP approach to image registration
- models on image and intensity pairs
  - Categorical / Multinomial
- MAP with known model parameters
- MAP with unknown model parameters
  - Dirichlet prior on parameters
  - Showed as special cases:
    - Registration by minimization of entropy of data
    - Registration by minimization of entropy of pooled data
    - MAP registration with known parameters (or Maximum Likelihood)
  - EM Algorithm to obtain estimates
    - Simple objective function: linear in counts
    - *Samson's Iteration*
- Experimental Results

The End