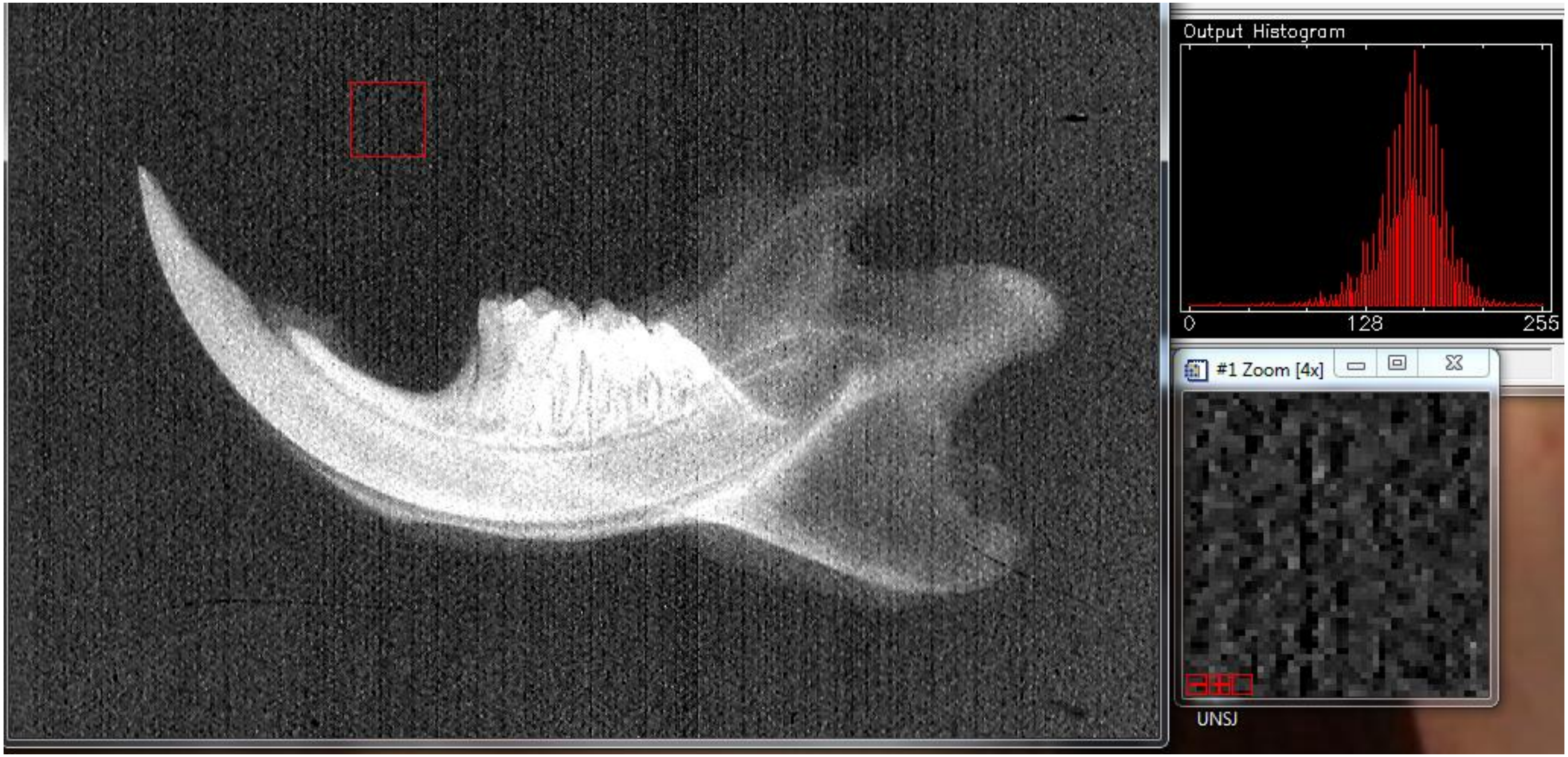
Comparison of Markovian Models for Supervised Image Segmentation

Josef Baumgartner
Georgina Flesia
Javier Gimenez
Jorge Martinez

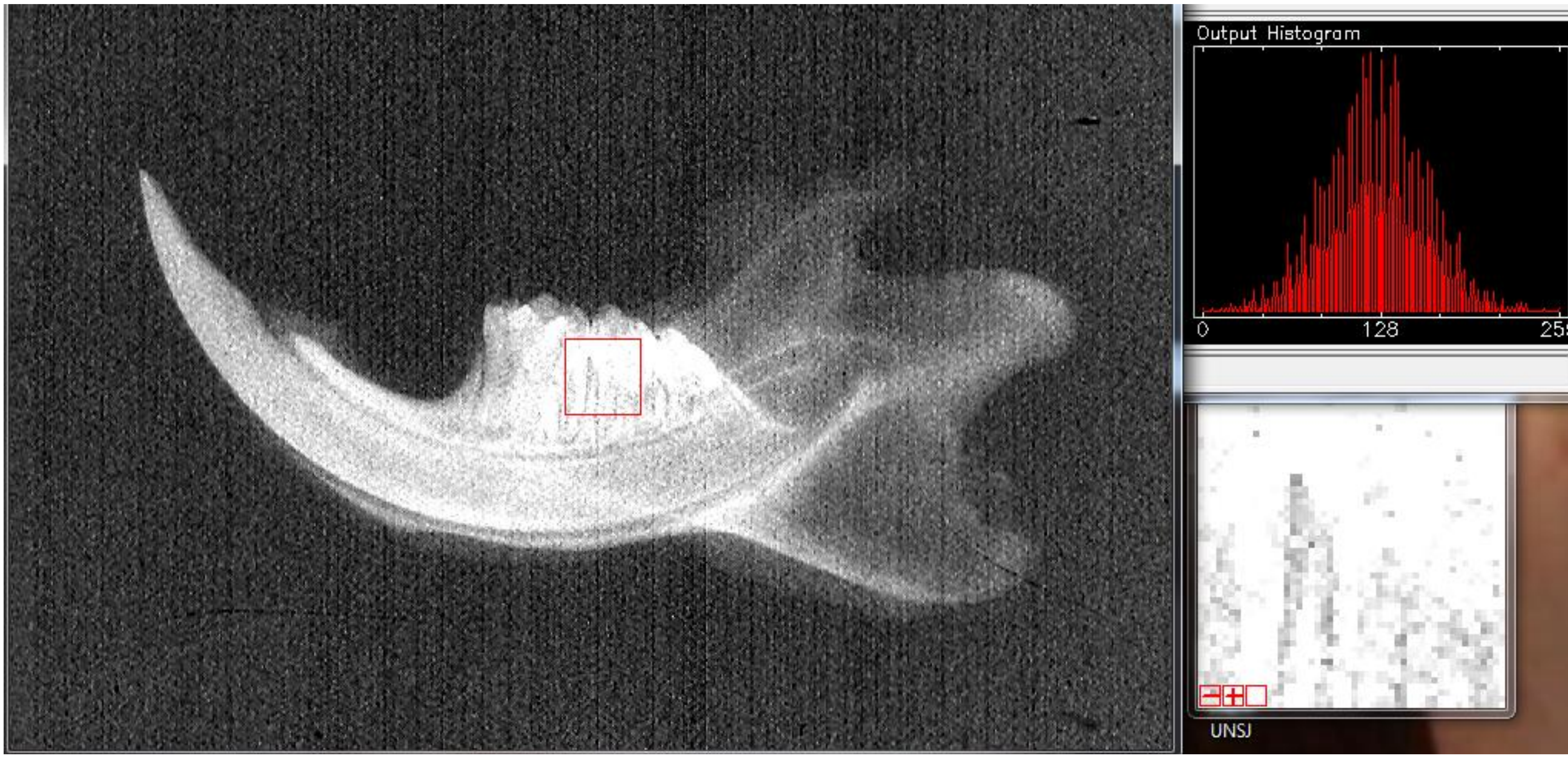
This talk explores three Gaussian Mixture Markov Random Field (GMMRF) based supervised image segmentation rules, induced by a variation of the Bayesian MAP principle. The rules are

- a 2dHMM (Hidden Markov Model) segmentation, where the hidden image class map is considered a 2nd order Markov Mesh, and the observations are modeled as a Gaussian mixture [1];
- a MAP-MRF rule which combines a Gaussian Markov Random field for the observations, and a Potts model for the a priori knowledge, estimating the solution via Besaj's Iterated conditional Modes,[2],
- the same GMMRF segmentation model, estimated using graph cut algorithms, [3].

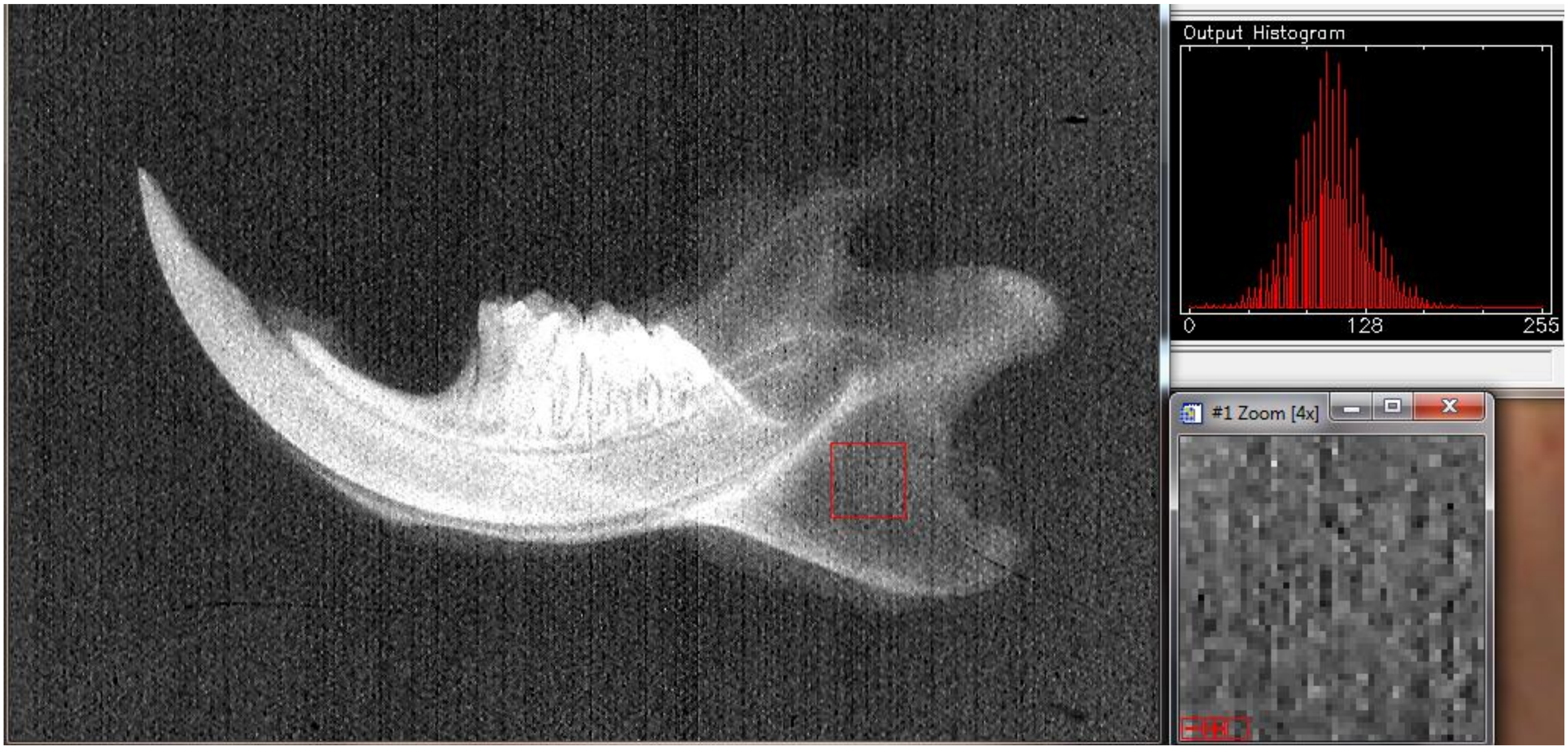
Data



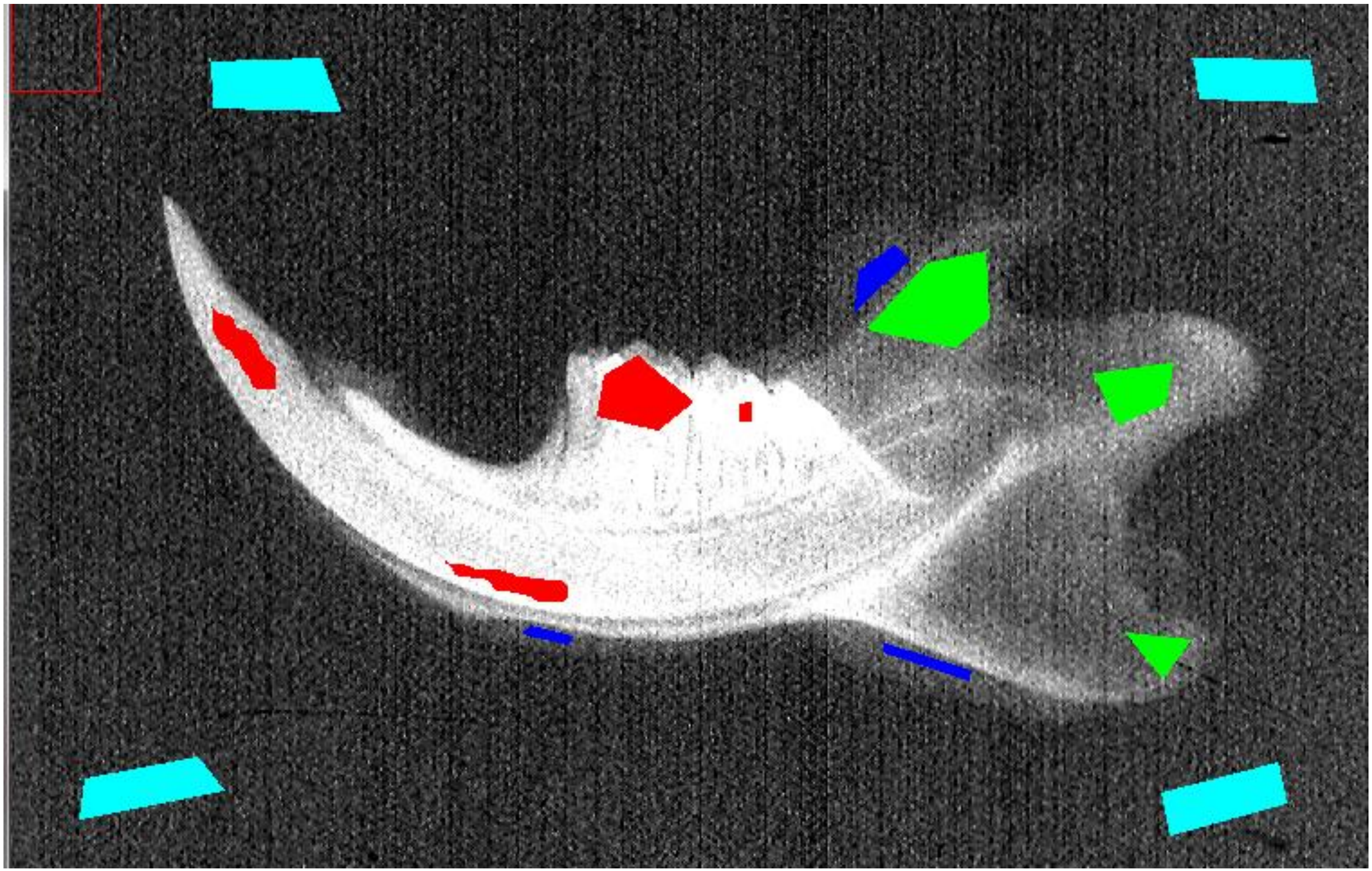
Data



Data

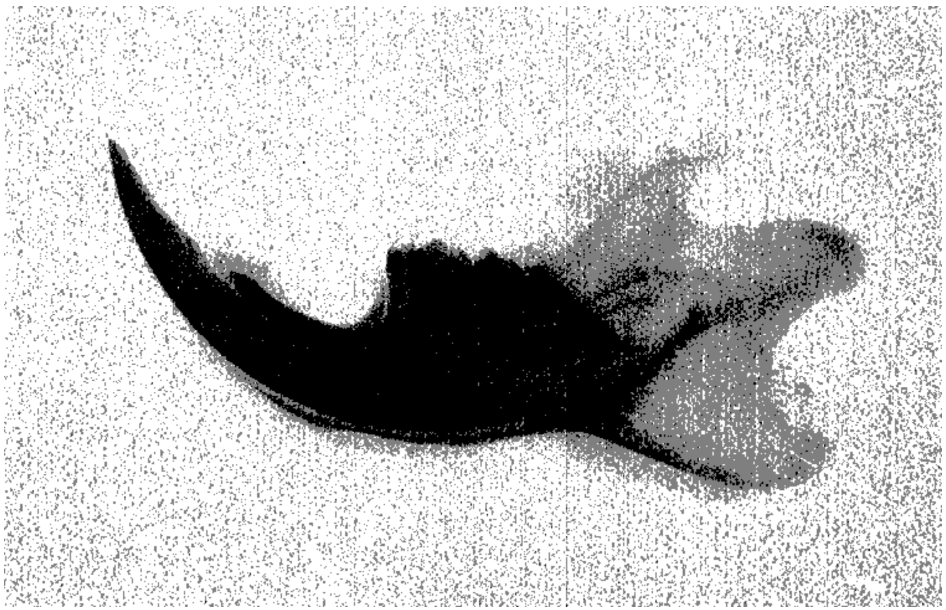


Training sets



Maximum likelihood

3 classes



4 classes

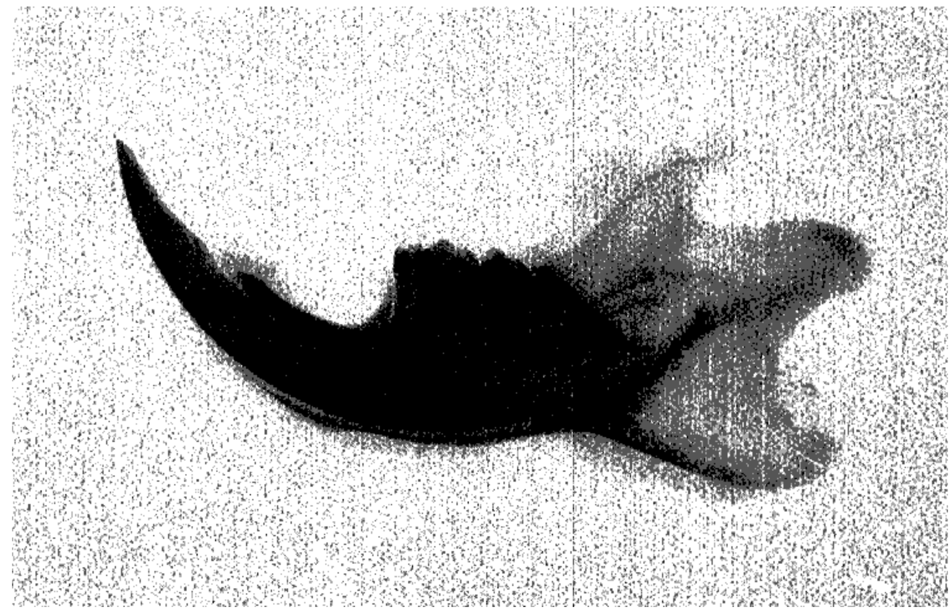
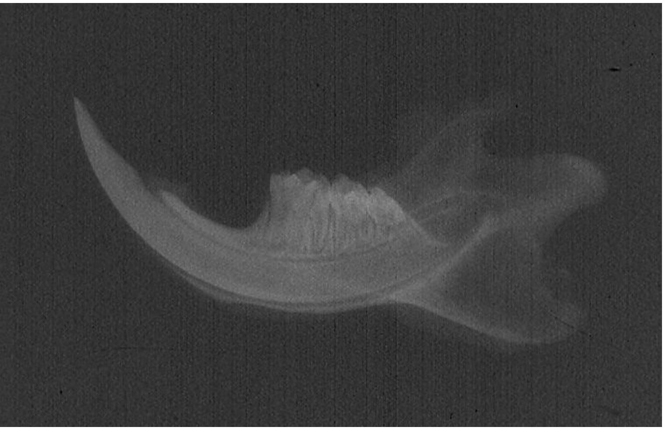
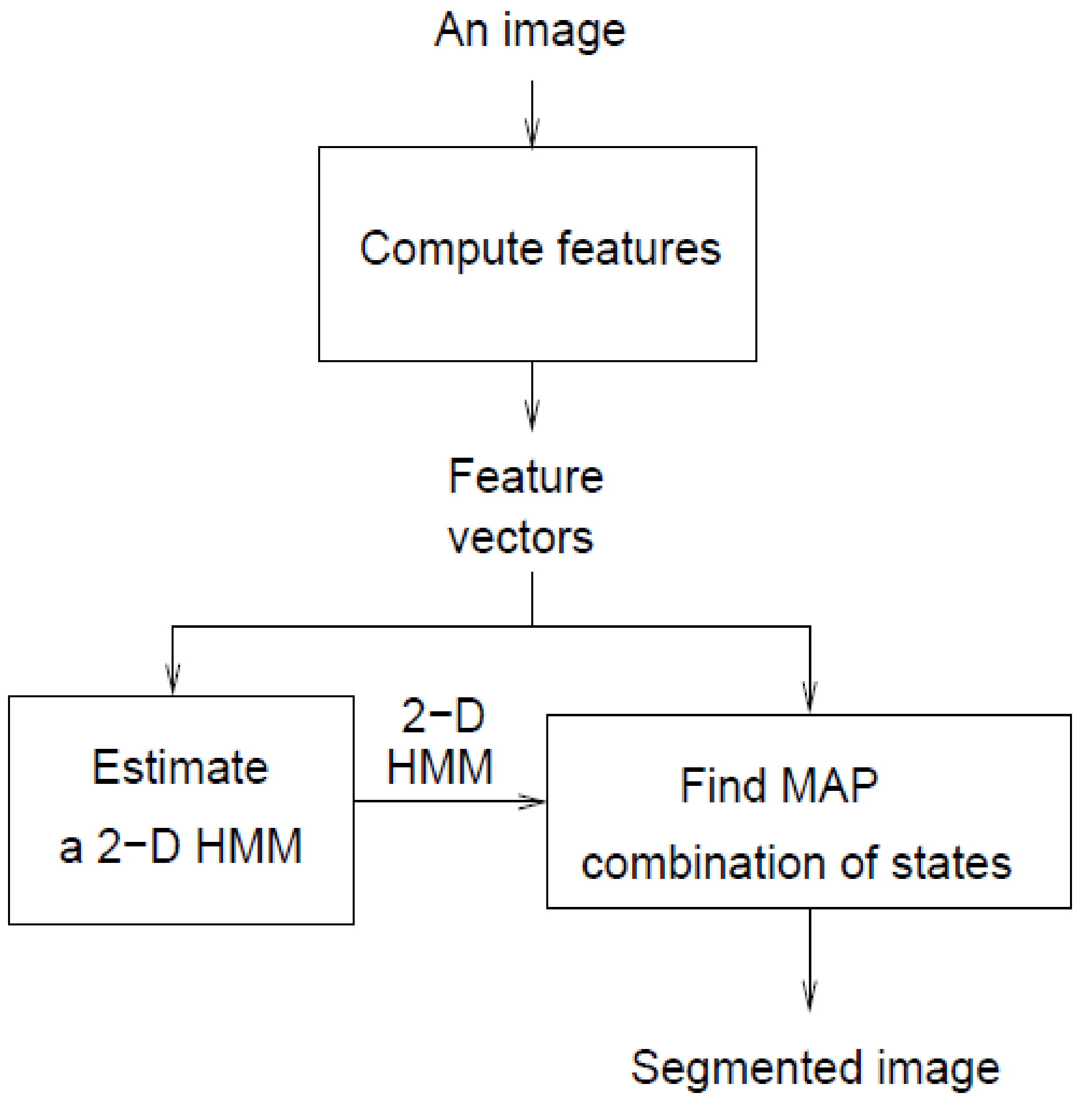


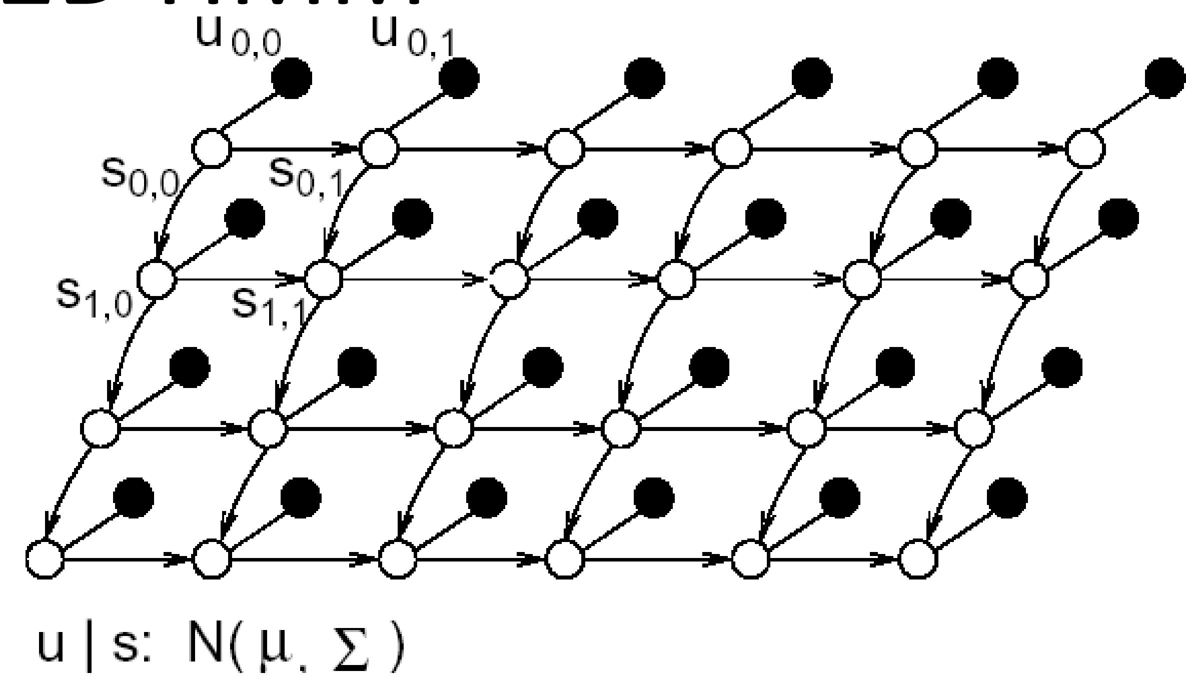
Image Segmentation





2D HMM

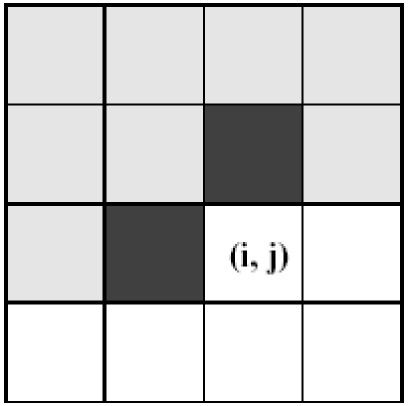
Regard an image as a grid. A feature vector is computed for each node.



- Each node exists in a hidden state.
- The states are governed by a Markov mesh (a causal Markov random field).
- Given the state, the feature vector is conditionally independent of other feature vectors and follows a normal distribution.
- The states are introduced to efficiently model the spatial dependence among feature vectors.
- The states are not observable, which makes estimation difficult.

2D HMM

The underlying states are governed by a Markov mesh.



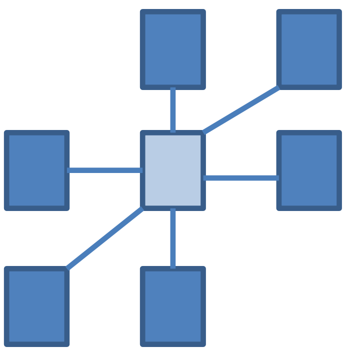
Partial order: (i',j')<(i,j) if i'<i or i'=i & j'<j

Context: the set of states for *(i', j'): (i', j')<(i, j)*

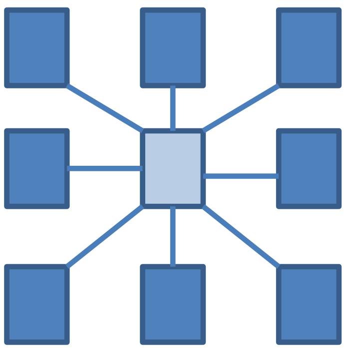
$P(s_{ij} \mid context)=a_{m,n,l}$, $m=s_{i-1,j}$, $n=s_{i,j-1}$ *and* $l=s_{i,j}$

Markov meshes are special cases of MRF

- Second and third order Markov meshes are MRF with neighborhoods:



a) 2nd order



b) 3rd order

Estimation of 2-D HMM

- Parameters to be estimated:
  - Transition probabilities
  - Mean and covariance matrix of each Gaussian distribution
- EM algorithm is applied for ML estimation.

EM Iteration

- Given the current model estimation $\phi^{(p)}$, the mean vectors and covariance matrices are updated by

$$\mu_m^{(p+1)} = \frac{\sum_{i,j} L_m^{(p)}(i,j) u_{i,j}}{\sum_{i,j} L_m^{(p)}(i,j)}$$

$$\boldsymbol{\Sigma}_m^{(p+1)} = \frac{\sum_{i,j} L_m^{(p)}(i,j)(u_{i,j} - \mu_m^{(p+1)})(u_{i,j} - \mu_m^{(p+1)})'}{\sum_{i,j} L_m^{(p)}(i,j)} .$$

- $L_m^{(p)}(i,j) = P\left(s_{i,j} = m \mid u_{i',j'}, c_{i',j'}, (i',j') \in \mathbb{N}; \phi^{(p)}\right)$
  - The probability of being in state m at block (i,j) given all the observed feature vectors, classes and model $\phi^{(p)}$.

Iteración EM

- The transition probabilities are updated as follows:

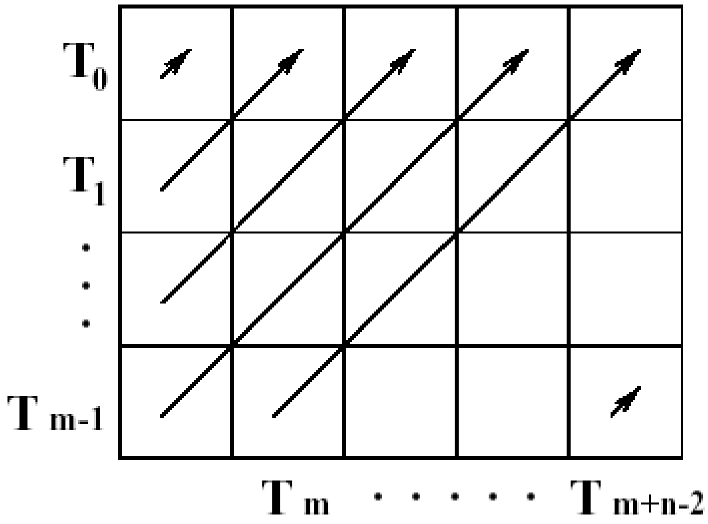
$$a_{m,n,l} = \frac{\sum_{i,j} H^{(p)}_{m,n,l}(i,j)}{\sum_{l'=1}^{M} \sum_{i,j} H^{(p)}_{m,n,l'}(i,j)}$$

- $H^{(p)}_{m,n,l}(i,j)$ is the probability of being in state m at block $(i-1,j)$, state n at block $(i,j-1)$ and state l at block (i,j) given the observed feature vectors, classes, and model $\phi^{(p)}$.

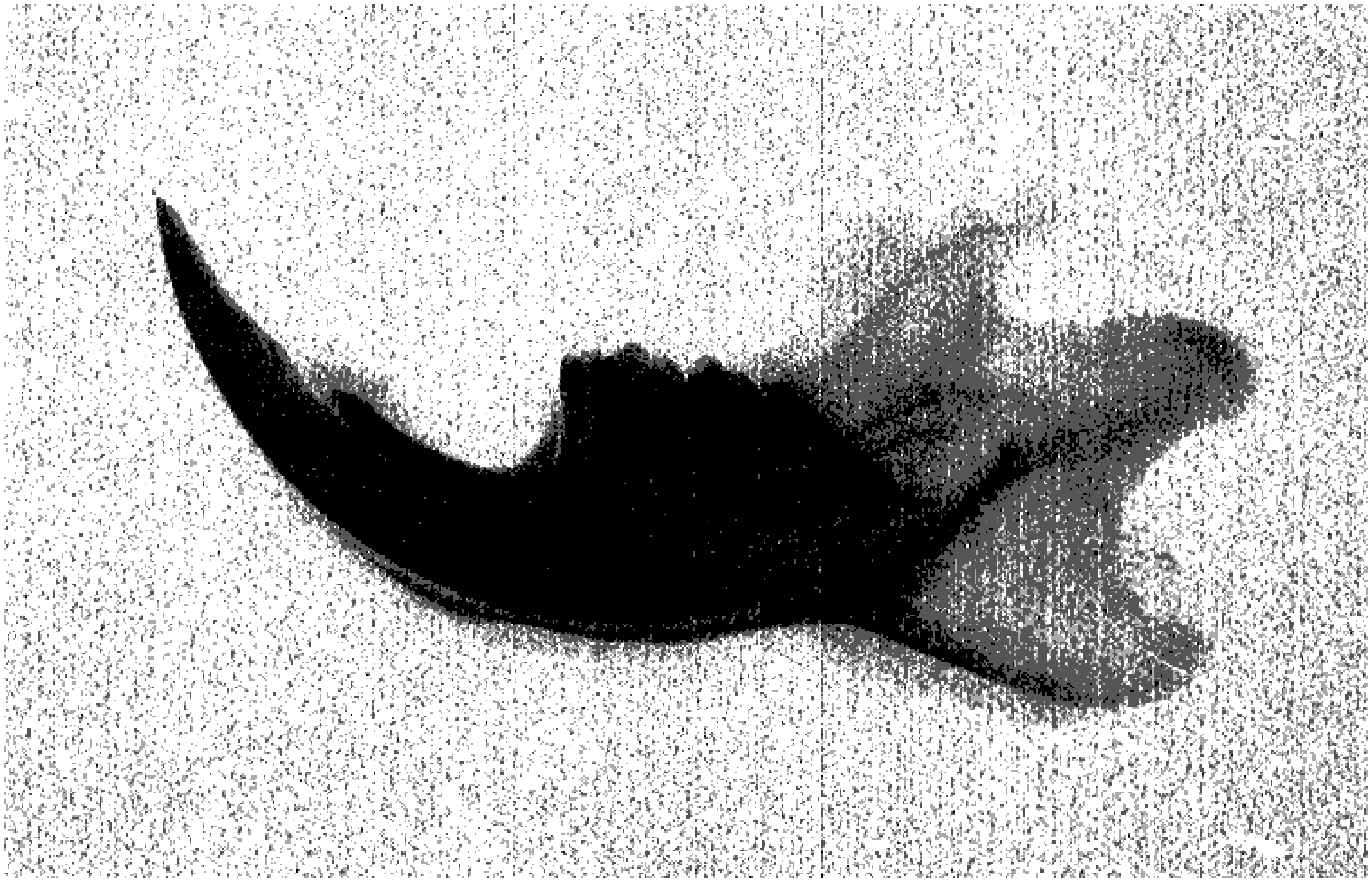
Computation Issues

- The brute force computation of $L_m^{(p)}(i,j)$ and $H_{m,n,l}^{(p)}(i,j)$: the order of computation is $w^2 M_0^{w^2}$.
- Forward-backward algorithm: reduce computation to the order of $w M_0^{2w}$.

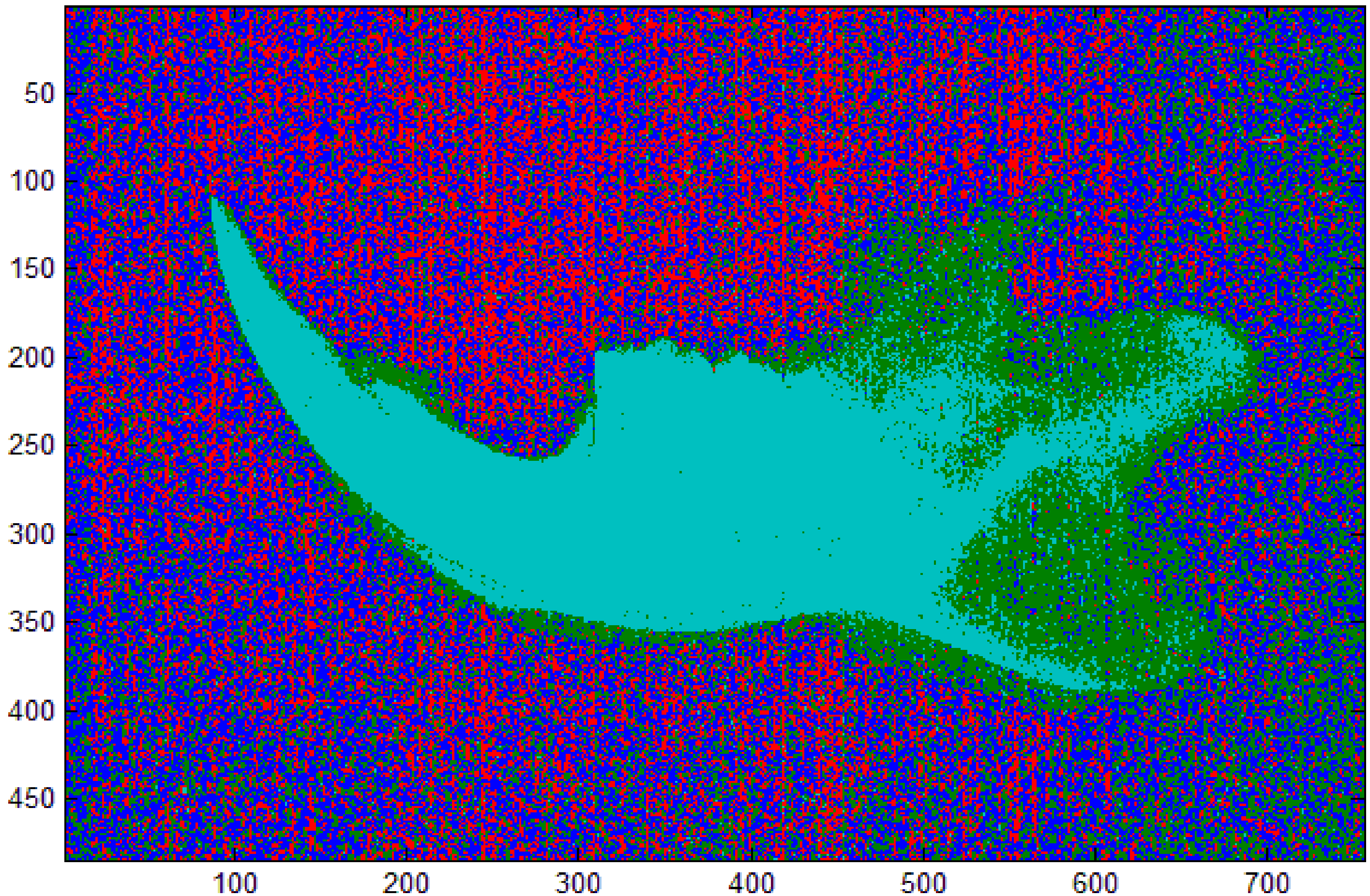
An approximation to the classification EM approach: constrained Viterbi



Maximum likelihood)



First results on 2d HMM



Gibbs Distributions

Potts strauss model

$$\Phi_A(c_A) = \begin{cases} -\beta & \text{si } A = \{s,t\} \text{ y } c_s = c_t \text{ , con } s \in V_t \\ 0 & \text{en otro caso} \end{cases}$$

$$H_\Lambda^\Phi(c) := \sum_{\substack{\{r,t\} \cap \Lambda \neq \emptyset \\ r \in V_t}} \Phi_{\{r,t\}}(\{c_r, c_t\})$$

$$\pi(c) \propto \exp\{-H_S^\Phi(c)\} = \exp\left\{-\sum_{\{\{r,t\}/r \in V_t\}} \Phi_{\{r,t\}}(\{c_r, c_t\})\right\} = \exp\{\beta M_c\}$$

Bayesian approach

$$\Pi(c|x) \propto p(x|c)\pi(c) \propto p(x|c)\exp\{\beta M_c\} = \prod_{s\in S} p(x_s|c_s)\exp\{\beta M_c\} = \exp\left\{\sum_{s\in S}\ln p(x_s|c_s) + \beta M_c\right\}$$

family of potentials

$$\Phi^x = \{\Phi^x_A\}_{\emptyset\subset A\subseteq S}$$

$$\Phi^x_A(c_A) = \begin{cases} -\ln p(x_s|c_s) & \text{si } A = \{s\}, s\in S \\ -\beta & \text{si } A = \{s,t\} \text{ y } c_s = c_t \text{ , con } s \in V_t \\ 0 & \text{en otro caso} \end{cases}$$

energy function

$$H_S^{\Phi^x}(c) = \sum_{A\subseteq S}\Phi_A(c_A) = -\sum_{s\in S}\ln p(x_s|c_s) - \sum_{\{\{r,t\}/r\in V_t\}}\Phi_{\{r,t\}}(\{c_r,c_t\}) = -\sum_{s\in S}\ln p(x_s|c_s) - \beta M_c.$$

we want to find c* such that

$$c^* = arg\max_{c\in L^S}\Pi(c|x)$$

which is equivalent to find a global maximum in the Gibbs distribution

Iterated Conditional Modes (ICM)

ICM proposes :

1) Inicial classification is a realization of the original distribution
2) a visitation scheme is selected
3) all pixels are visited in turn , and their labels are updated for the label that maximizes

$$\exp\{\ln p(x_s|l) + \beta N_s^c(l)\}$$

$$\ln p(x_s|l) + \beta N_s^c(l)$$

4) An iteration number is fixed

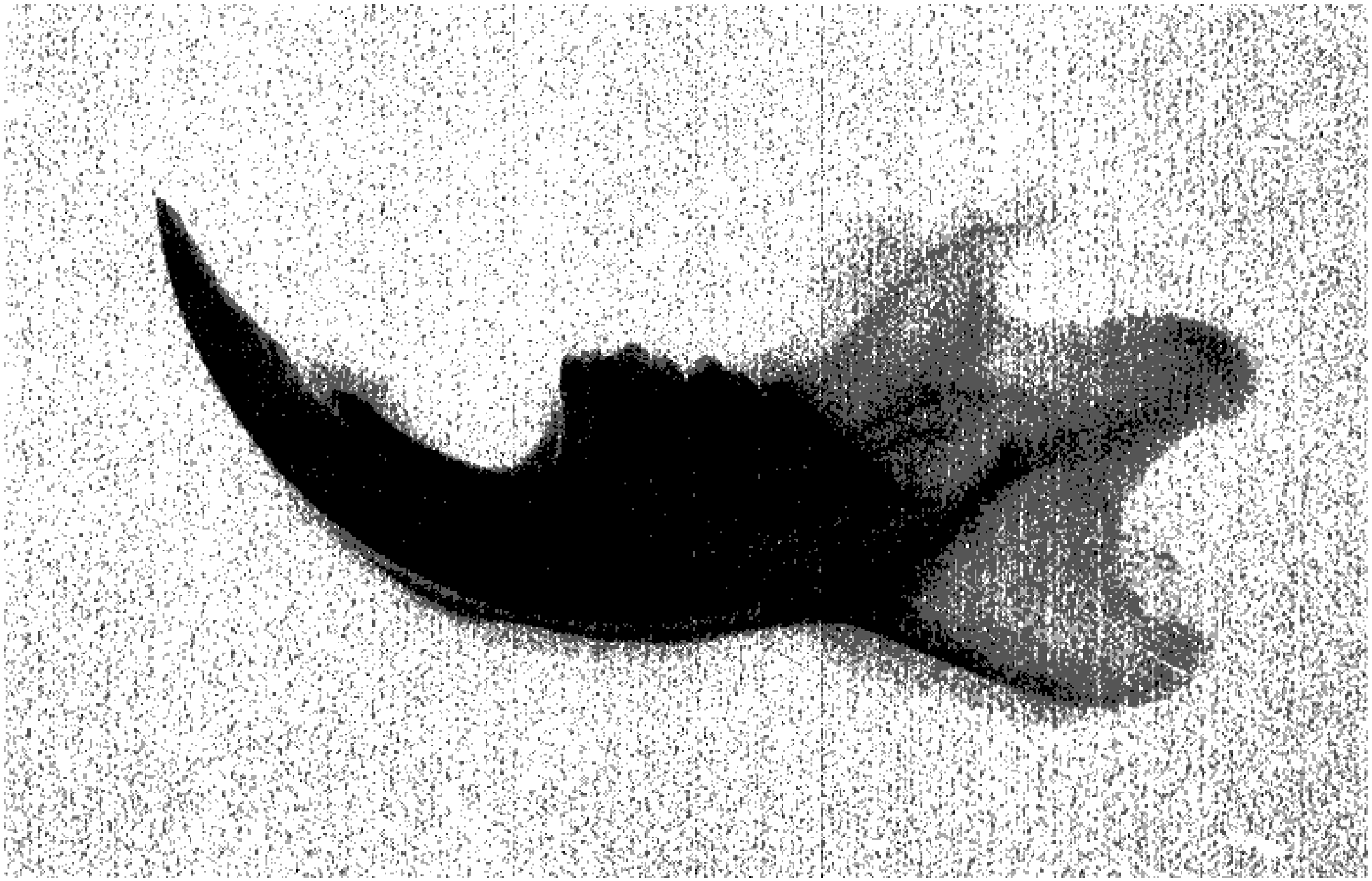
Iterated Conditional Modes (ICM)

Working over the last expresion , at each site s, the label is changed by the one that maximizes

$$g(l) = \frac{1}{2}\left(-\ln(det(M_l)) - (x_s - \mu_l)^T M_l^{-1}(x_s - \mu_l)\right) + \beta \#\{t \in V_s / c_t = l\},$$

The first term is the same as the Gaussian ML classifier, the second term is the contextual component , and if β >0 gives weight to the labels around the site s.

Maximum Likelihood



Iterated Conditional Modes (ICM)

ICM - $\beta = 3$ - 10 iteraciones

Iterated Conditional Modes (ICM)

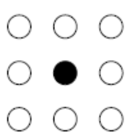
ICM - $\beta = 10$ - 10 iteraciones

Pseudo-likelihood equations for β

$$\sum_{s \in S} N_s^c(c_s) - \sum_{s \in S} \frac{\sum_{l \in L} N_s^c(l) \exp\{\beta N_s^c(l)\}}{\sum_{l \in L} \exp\{\beta N_s^c(l)\}} = 0$$

Pseudo-likelihood equations for β

potts model



$$
\begin{aligned}
\frac{\partial}{\partial\beta}\log PL(\beta) = & \sum_{s\in\Omega} U_s(m_s) - \frac{8e^{8\hat\beta}}{e^{8\hat\beta}+M-1}K_1 - \frac{7e^{7\hat\beta}+e^{\hat\beta}}{e^{7\hat\beta}+e^{\hat\beta}+M-2}K_2 - \frac{6e^{6\hat\beta}+2e^{2\hat\beta}}{e^{6\hat\beta}+e^{2\hat\beta}+M-2}K_3 \\
& - \frac{6e^{6\hat\beta}+2e^{\hat\beta}}{e^{6\hat\beta}+2e^{\hat\beta}+M-3}K_4 - \frac{5e^{5\hat\beta}+3e^{3\hat\beta}}{e^{5\hat\beta}+e^{3\hat\beta}+M-2}K_5 - \frac{5e^{5\hat\beta}+2e^{2\hat\beta}+e^{\hat\beta}}{e^{5\hat\beta}+e^{2\hat\beta}+e^{\hat\beta}+M-3}K_6 \\
& - \frac{5e^{5\hat\beta}+3e^{\hat\beta}}{e^{5\hat\beta}+3e^{\hat\beta}+M-4}K_7 - \frac{8e^{4\hat\beta}}{2e^{4\hat\beta}+M-2}K_8 - \frac{4e^{4\hat\beta}+3e^{3\hat\beta}+e^{\hat\beta}}{e^{4\hat\beta}+e^{3\hat\beta}+e^{\hat\beta}+M-3}K_9 \\
& - \frac{4e^{4\hat\beta}+4e^{2\hat\beta}}{e^{4\hat\beta}+2e^{2\hat\beta}+M-3}K_{10} - \frac{4e^{4\hat\beta}+2e^{2\hat\beta}+2e^{\hat\beta}}{e^{4\hat\beta}+e^{2\hat\beta}+2e^{\hat\beta}+M-4}K_{11} \\
& - \frac{4e^{4\hat\beta}+4e^{\hat\beta}}{e^{4\hat\beta}+4e^{\hat\beta}+M-5}K_{12} - \frac{6e^{3\hat\beta}+2e^{2\hat\beta}}{2e^{3\hat\beta}+e^{2\hat\beta}+M-3}K_{13} - \frac{6e^{3\hat\beta}+2e^{\hat\beta}}{2e^{3\hat\beta}+2e^{\hat\beta}+M-4}K_{14} \\
& - \frac{3e^{3\hat\beta}+4e^{2\hat\beta}+e^{\hat\beta}}{e^{3\hat\beta}+2e^{2\hat\beta}+e^{\hat\beta}+M-4}K_{15} - \frac{3e^{3\hat\beta}+2e^{2\hat\beta}+3e^{\hat\beta}}{e^{3\hat\beta}+e^{2\hat\beta}+3e^{\hat\beta}+M-5}K_{16} \\
& - \frac{3e^{3\hat\beta}+5e^{\hat\beta}}{e^{3\hat\beta}+5e^{\hat\beta}+M-6}K_{17} - \frac{8e^{2\hat\beta}}{4e^{2\hat\beta}+M-4}K_{18} - \frac{6e^{2\hat\beta}+2e^{\hat\beta}}{3e^{2\hat\beta}+2e^{\hat\beta}+M-5}K_{19} \\
& - \frac{4e^{2\hat\beta}+4e^{\hat\beta}}{2e^{2\hat\beta}+4e^{\hat\beta}+M-6}K_{20} - \frac{2e^{2\hat\beta}+6e^{\hat\beta}}{e^{2\hat\beta}+6e^{\hat\beta}+M-7}K_{21} - \frac{8e^{\hat\beta}}{8e^{\hat\beta}+M-8}K_{22} = 0
\end{aligned}
$$

$$\Pi(c|x) \propto p(x|c)\pi(c) \propto p(x|c)\exp\{\beta M_c\} = \prod_{s\in S} p(x_s|c_s)\exp\{\beta M_c\} = \exp\left\{\sum_{s\in S} \ln p(x_s|c_s) + \beta M_c\right\}$$

$$c^* = arg\max_{c\in L^S} \Pi(c|x)$$

$$\pi(f) = Z^{-1} exp\left(-\sum_{\{p,q\}\in\mathcal{N}} V_{\{p,q\}}(f_p, f_q)\right).$$

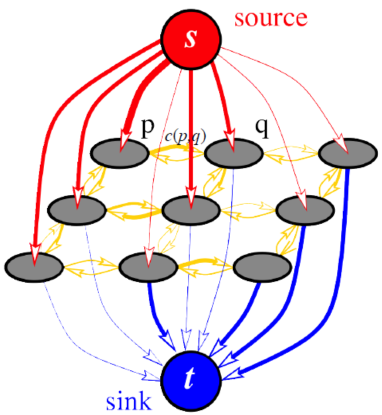
Writing the Potts ecuations as general energy functions, we can use the energy minimization methods to solve the labeling problem

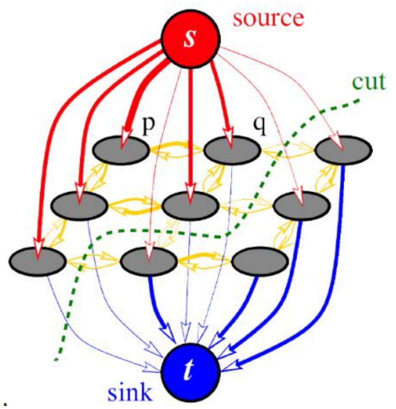
$$f^* = argmax_{f\in\mathcal{F}} exp\left(-\sum_{\{p,q\}\in\mathcal{N}} V_{\{p,q\}}(f_p, f_q) - \sum_{p\in\mathcal{P}} D_p(f_p)\right)$$

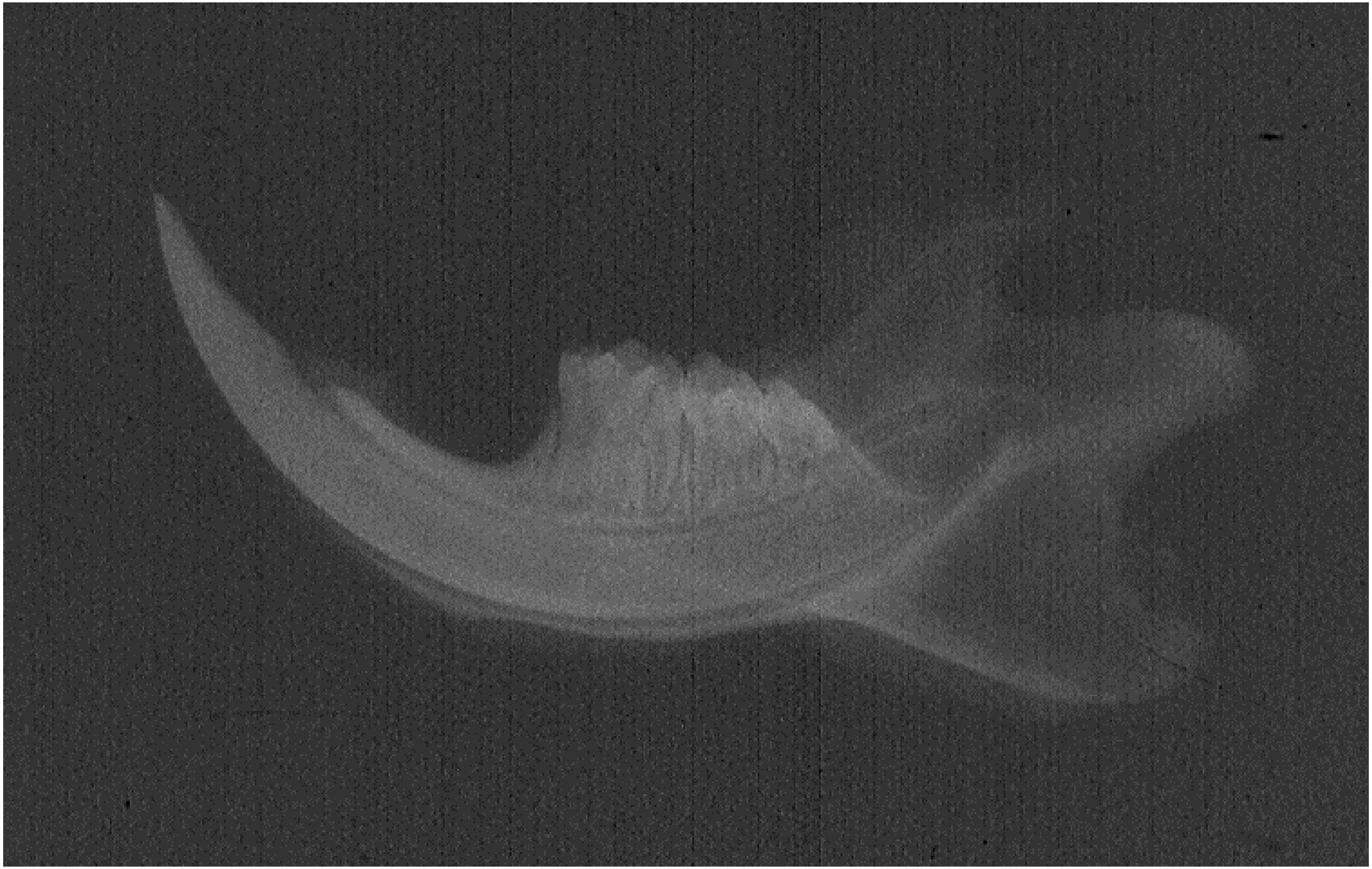
MAP problem is equivalent to minimize

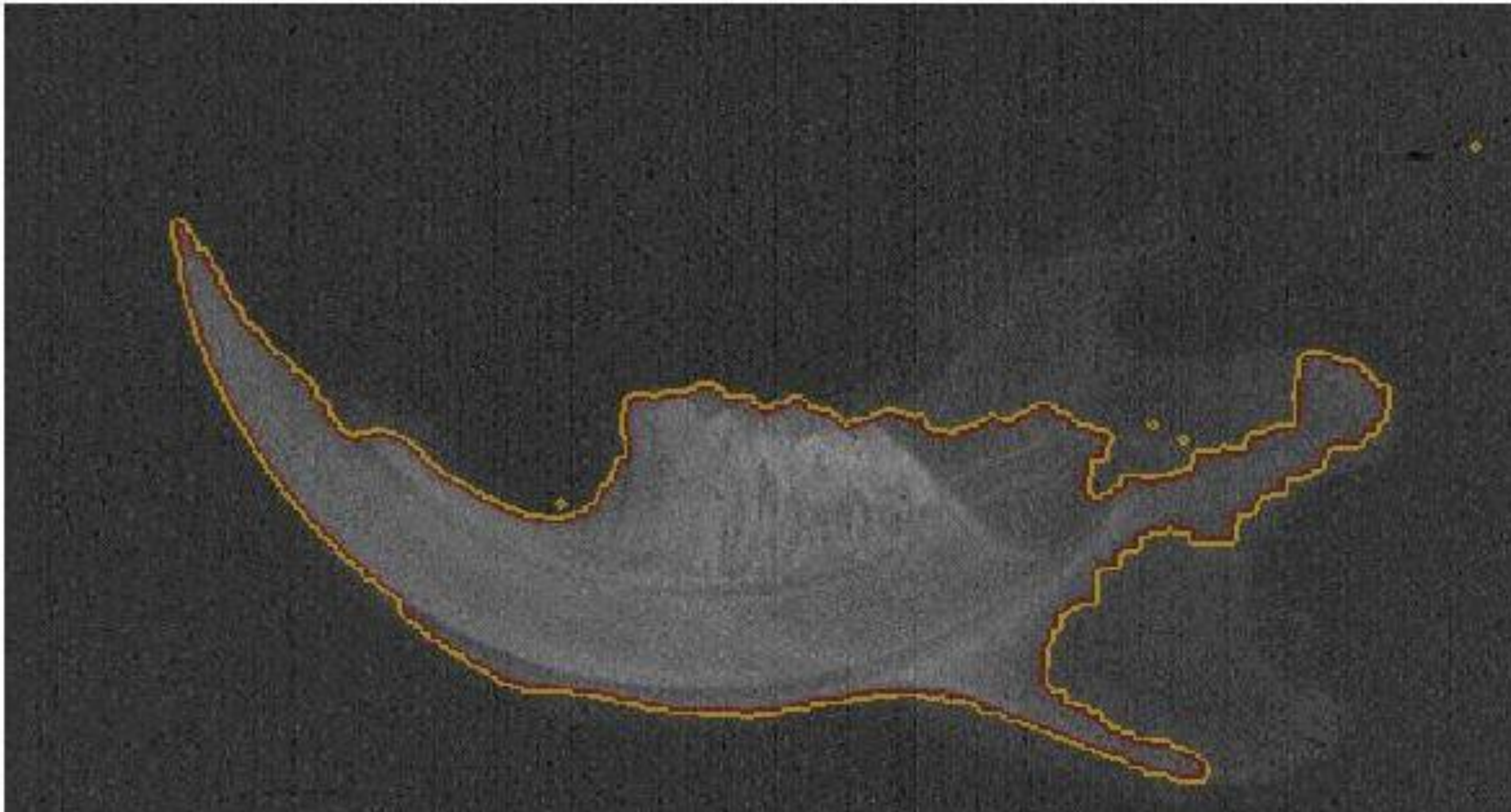
$$E(f) = \underbrace{\sum_{\{p,q\}\in\mathcal{N}} V_{\{p,q\}}(f_p, f_q)}_{\text{Prior knowledge: Smoothness term}} + \underbrace{\sum_{p\in\mathcal{P}} D_p(f_p)}_{\text{Local beliefs: Data term}}$$

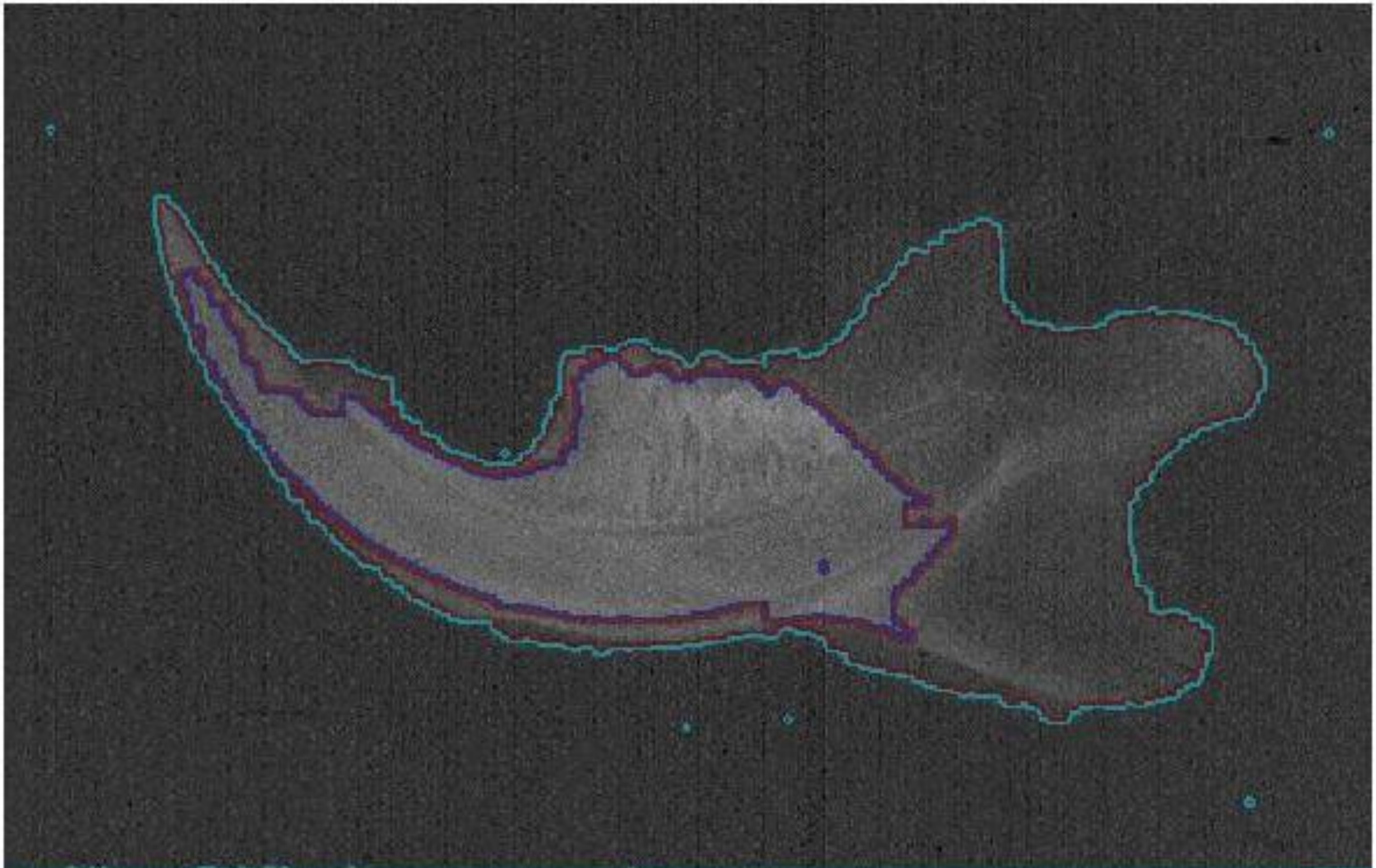
max flow min cut algorithms are efficents ways of solving the problem

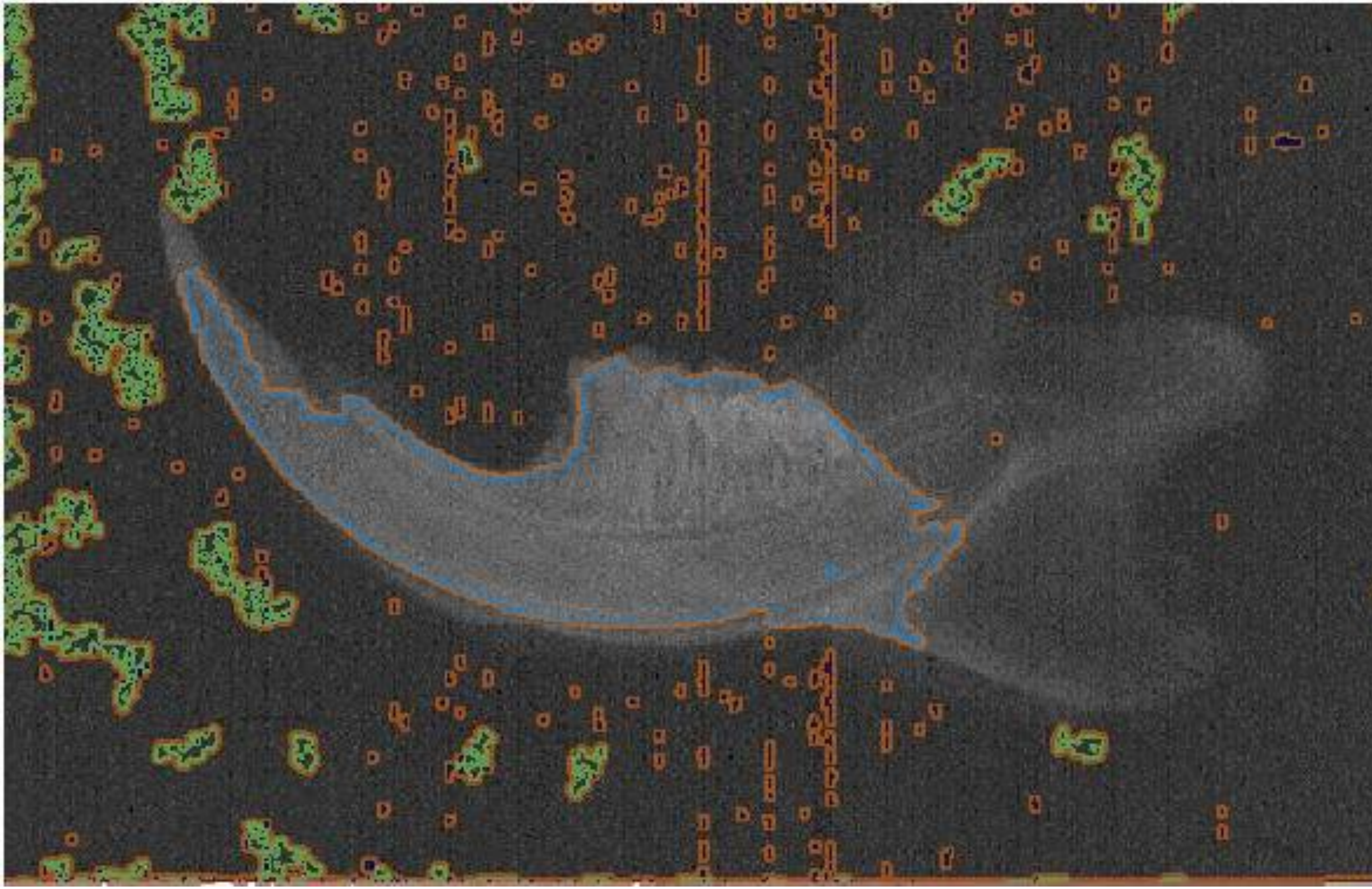












Quantitative performance were assessed within a hypothesis test based on Cohen's Kappa coefficient, estimating classification errors and confusion matrix by 10-fold cross validation

$$K = \frac{N \sum_i c_{ii} - \sum_i x_{i+} x_{+i}}{N^2 - \sum_i x_{i+} x_{+i}}$$

showing the obvious fact that MV is improved by contextual classification.

Differences between methods were also observed, as well as differences in the performance of the methods when different initializations were used

All algorithms, using the learned parameters, generates good object-segmentations with little interaction, besides selection of training data. However, sub-optimal learning proves to be frail, which limits the complexity of usable models, and hence also the achievable error rate.

Conclusions

- Kappa is a very old measure
- It can not say when background errors are gross than edge errors
- Segmentations may be visually very differents but similar by Kappa
- Segementations may be visually different and different by Kappa but there is no way to tell which is better
- There is some research to be done in quality measures yet.

References

[1] Xiang Ma, D. Schonfeld and A. Khokhar. A general two dimensional hidden Markov Model and its application in image classication. IEEE International Conference on Image Processing, 2007. ICIP 2007. Sept. 16 2007-Oct. 19 2007. San Antonio, TX, Volume: 6 :VI - 41 - VI - 44.

[2] A. Levada, N. Mascarenhas, A. Tannus. A novel MAP-MRF approach for multispectral image contextual classification using a combination of suboptimal iterative algorithms Pattern Recognition Letters:31 (2010): 1795--1808.

[3] A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr. Interactive image segmentation using an adaptive GMMRF model. Proc. Eur. Conf. on Computer Vision, ECCV (2004).