Entropic-graphs for pattern matching

Alfred O. Hero

Dept. EECS, Dept Biomed. Eng., Dept. Statistics

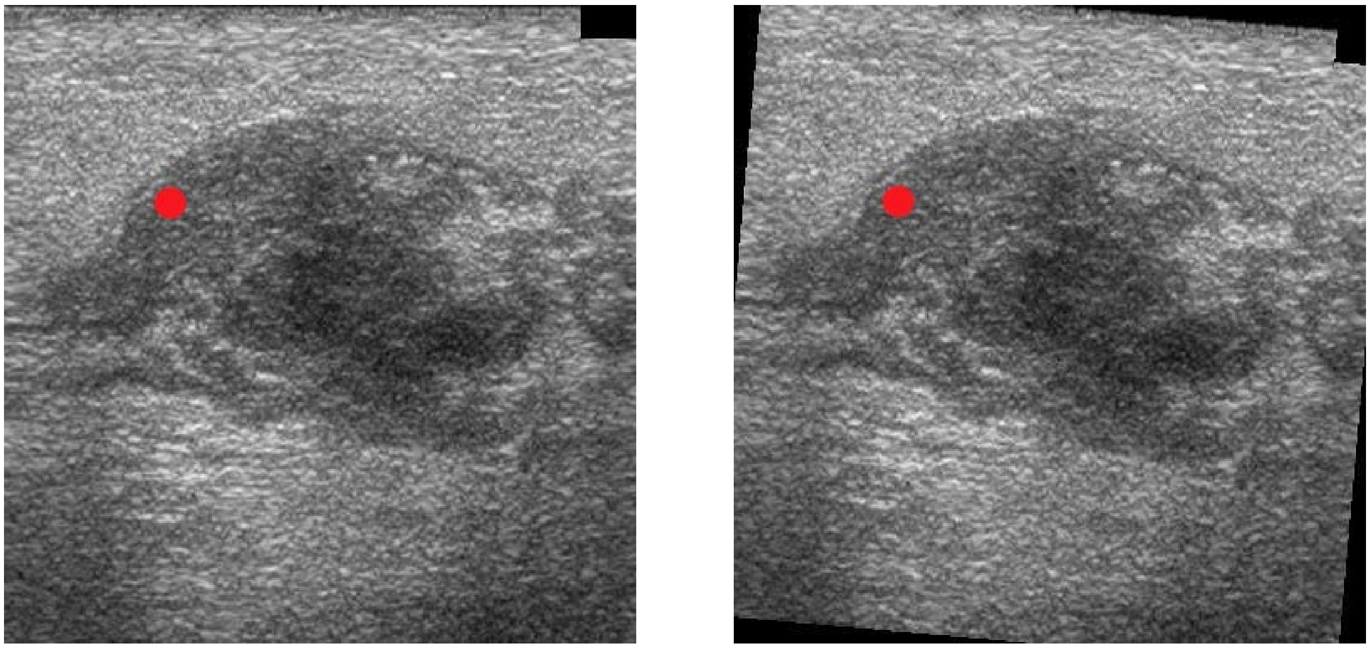
University of Michigan - Ann Arbor

hero@eecs.umich.edu

http://www.eecs.umich.edu/˜hero

Collaborators: Huzefa Heemuchwala, Jose Costa, Bing Ma, Olivier Michel

Motivation: Image Registraion



(a) Image I^R (b) Image I^T

Figure 1: Single Pixel Coincidences (Left and right: reference image I^R at 0^o and rotated image I^T at 8^o)

Single-Pixel Scatterplot $(Z_j^R, Z_j^T)_{j=1}^p$

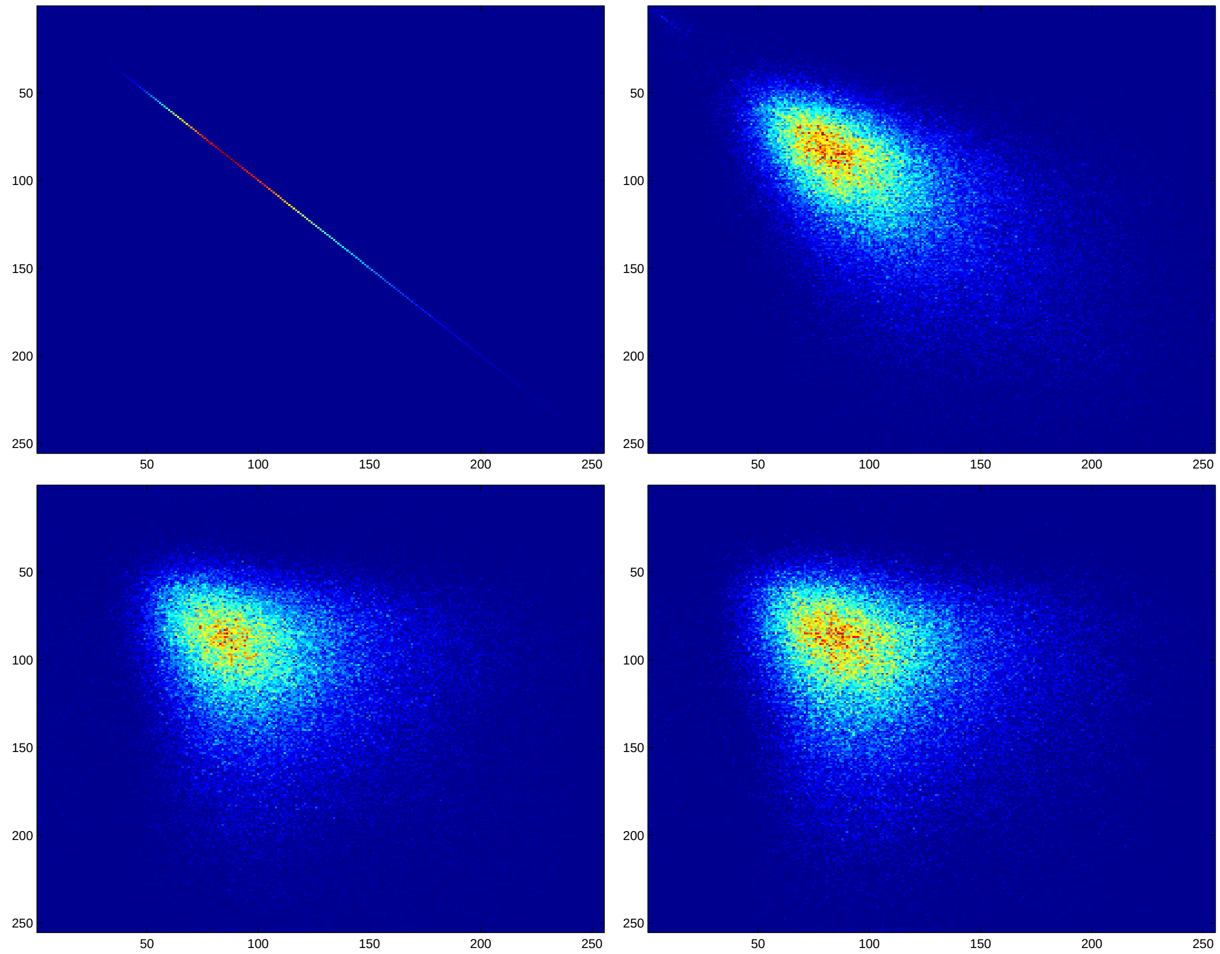


Figure 2: Grey level scatterplots. 1st Col: target=reference slice. 2nd Col: target = reference+1 slice.

Renyi α-Divergence

The Rényi α-divergence of fractional order $\alpha \in [0,1]$ [Rényi:61,70]

$$D_\alpha(f_1 \parallel f_0) \quad = \quad \frac{1}{\alpha - 1} \ln \int f_1 \left(\frac{f_1}{f_0}\right)^\alpha dx = \frac{1}{\alpha - 1} \ln \int f_1^\alpha f_0^{1-\alpha} dx$$

- α-Divergence vs α-Entropy

$$H_\alpha(f_1) = \frac{1}{1-\alpha} \ln \int f_1^\alpha dx = -D_\alpha(f_1 \parallel f_0)|_{f_0 = U([0,1]^d)}$$

- α-Divergence vs. Batthacharyya-Hellinger distance

$$D_{BH}^2(f_1 \parallel f_0) \quad = \quad \int \left(\sqrt{f_1} - \sqrt{f_0}\right)^2 dx = 2\left(1 - \exp\left(\tfrac{1}{2} D_{\frac{1}{2}}(f_1 \parallel f_0)\right)\right)$$

- α-Divergence vs. Kullback-Liebler divergence (Shannon MI)

$$\lim_{\alpha \to 1} D_\alpha(f_1 \| f_0) = \int f_1 \ln \frac{f_1}{f_0} dx.$$

Rényi α-divergence and Error Exponents

Observe i.i.d. sample $\underline{W} = [W_1, \ldots, W_n]$

$$
\begin{aligned}
H_0 &: \quad W_j \sim f_0(w) \\
H_1 &: \quad W_j \sim f_1(w)
\end{aligned}
$$

Bayes probability of error

$$
P_e(n) \quad = \quad \beta(n)P(H_1) + \alpha(n)P(H_0)
$$

LDP gives Chernoff bound [Dembo&Zeitouni:98]

$$
\liminf_{n\to\infty} \frac{1}{n} \log P_e(n) = -\sup_{\alpha\in[0,1]} \{(1-\alpha)D_\alpha(f_1 \| f_0)\}.
$$

Special case: image registration

$$
H_0 \;:\; \{Z^R, Z^T\} \text{ independent}, \quad \text{vs.} \quad H_1 \;:\; \{Z^R, Z^T\} \text{ dependent}
$$

Error exponent is α-MI.

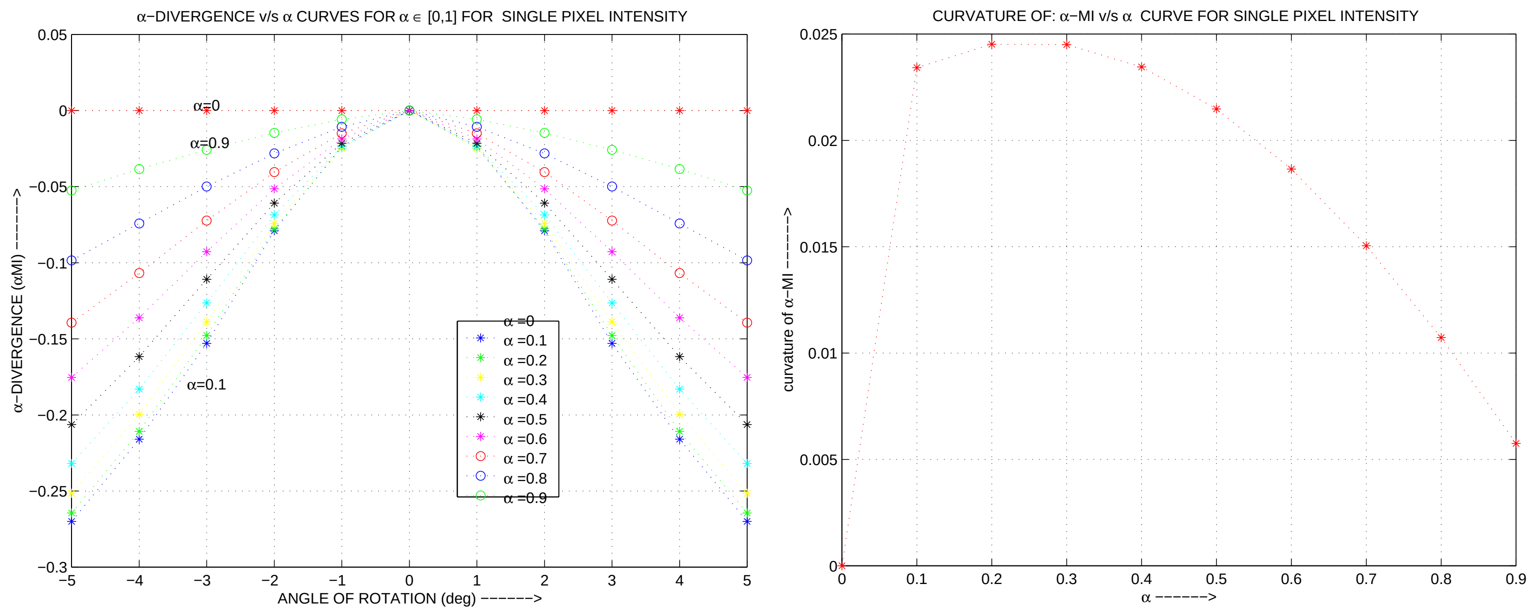


Figure 3: Left: α-Divergence as function of angle. Right: Resolution of α-Divergence as function of alpha

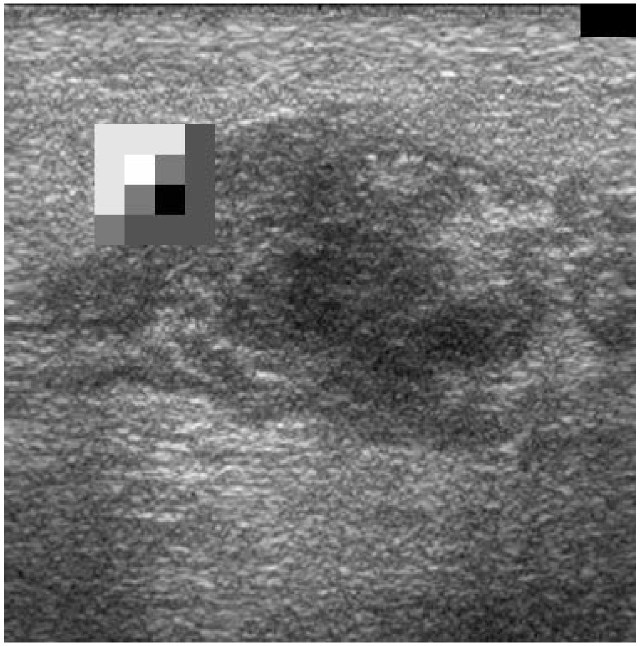
Higher Level Features

Disadvantages of single-pixel features:

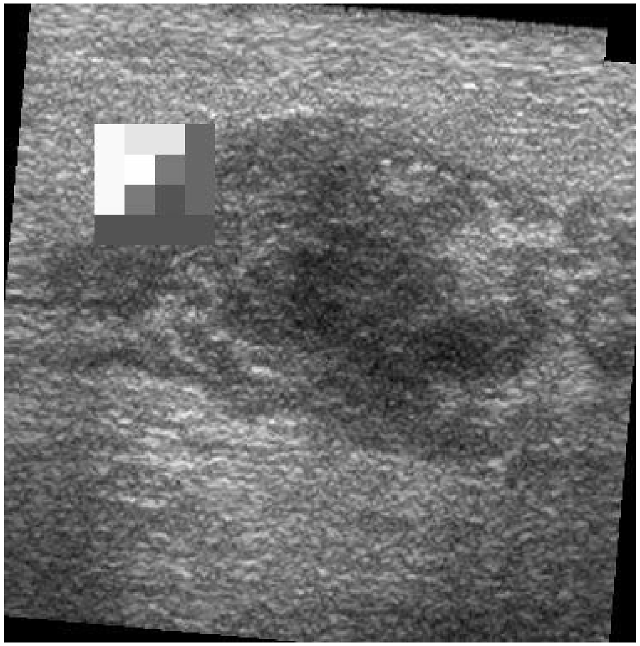
- Only depends on histogram of single pixel pairs
- Insensitive to spatial reording of pixels in each image
- Difficult to select out grey level anomalies (shadows, speckle)
- Spatial discriminants fall outside of single pixel domain

- **Alternative**: Aggregate spatial features

Local Tags



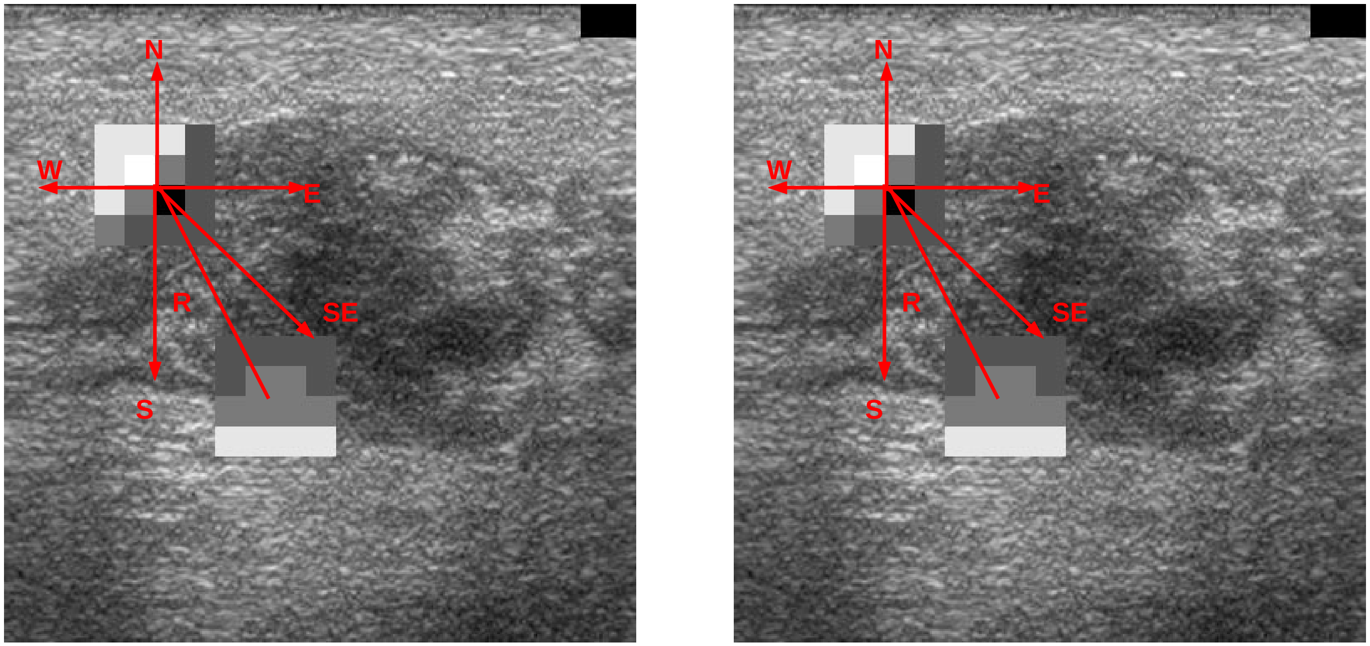
(a) Image I^R



(b) Image I^T

Figure 4: Local Tag Coincidences

Spatial Relations Between Local Tags



(a) Image I^R (b) Image I^T

Figure 5: Spatial Relation Coincidences

US Registration Comparisons

	151	142	162
pixel	0.6/0.9	0.6/0.3	0.6/0.3
tag	0.5/3.6	0.5/3.8	0.4/1.4
spatial-tag	0.99/14.6	0.99/8.4	0.6/8.3

Table 1: Numerator =optimal values of α and Denominator = maximum resolution of mutual α-information for registering various images (Cases 151, 142, 162) using various features (pixel, tag, spatial-tag, ICA).

Alternative: Projection-coefficient vector wrt ICA basis

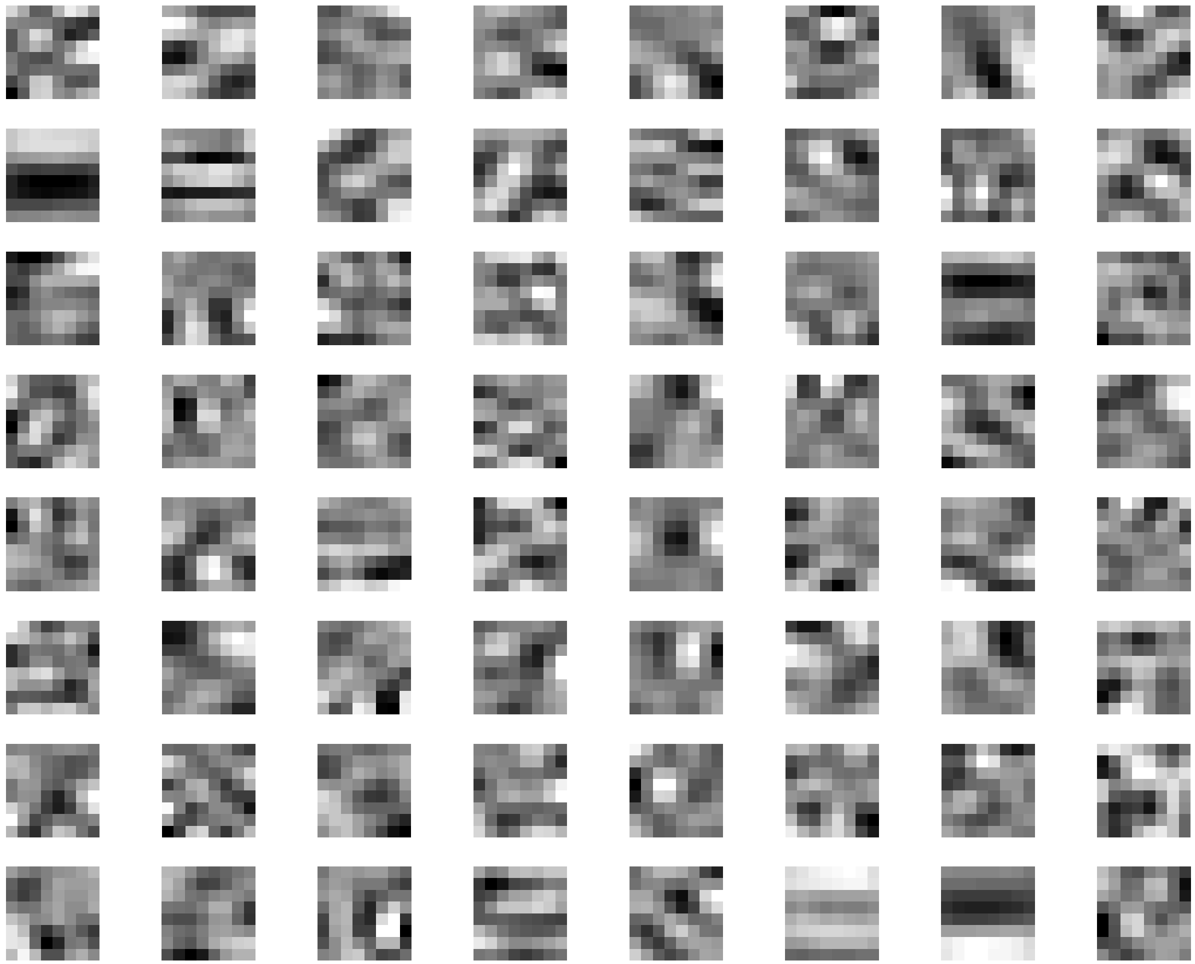


Figure 6: *Estimated ICA basis set for ultrasound breast image database*

Feature-based Indexing: Challenges

- How to best select discriminating features?
  - *Require training database of images to learn feature set*
  - Apply cross-validation...
  - ...bagging, boosting, or randomized selection?
- How to compute α-MI for multi-dimensional features?
  - *Tag space is of high cardinality:* $256^{16} \geq 10^{32}$
  - *ICA projection-coefficient space is multi-dimensional continuum*
  - Soln 1: partition feature space and count coincidences...
  - Soln 2: apply density estimation and ...
  - ... plug into the α-MI
  - Soln 3: estimate α-MI directly

Methods of Entropy/Divergence Estimation

- $Z = (Z^R, Z^T)$: a statistic (feature pair)
- $\{Z_i\}$: n i.i.d. realizations from $f(Z)$

Objective: Estimate

$$H_\alpha(f) = \frac{1}{1-\alpha} \ln \int f^\alpha(x) dx$$

1. Parametric density estimation methods
2. Non-parametric density estimation “plug-in” methods
3. Non-parametric minimal-graph estimation methods

Minimal Graphs: Minimal Spanning Tree (MST)

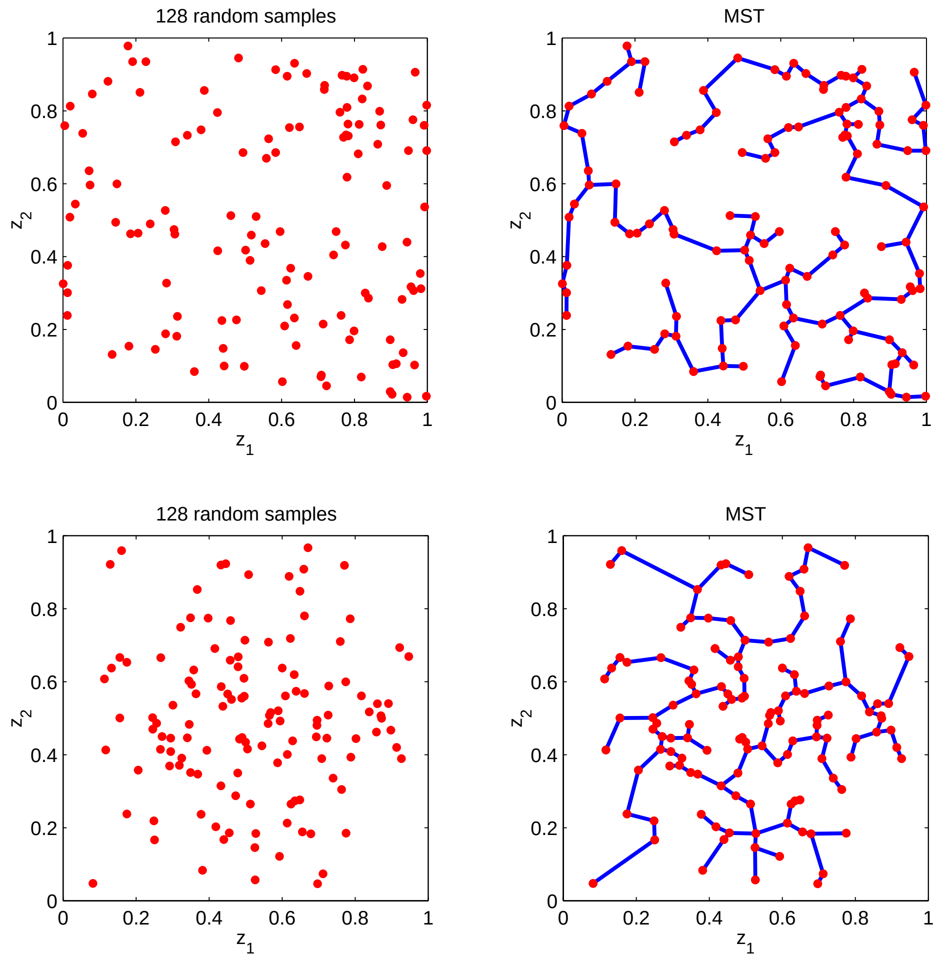


Figure 7:

Convergence of MST

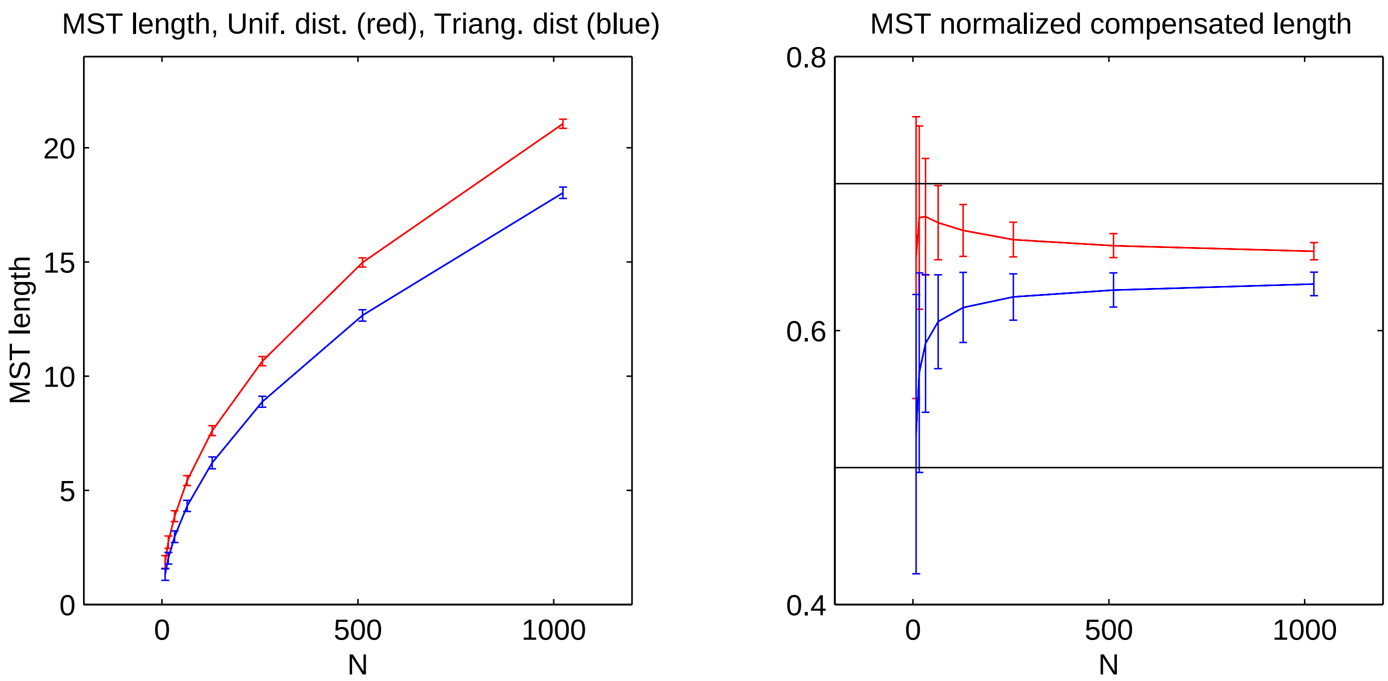


Figure 8:

Asymptotics: the BHH Theorem and entropy estimation

Define the MST length functional

$$L_\gamma(X_n) = \min_{T_n} \sum_{e \in T_n} \|e\|^\gamma.$$

Theorem 1

(Beardwood&etal:Camb59,Steele:95,Redmond&Yukich:SPA96)

$$\lim_{n\to\infty} L_\gamma(X_n)/n^{(d-\gamma)/d} \quad = \quad \beta_{L_\gamma,d} \int f(x)^{(d-\gamma)/d} dx, \qquad (a.s.)$$

Or, letting $\alpha = (d-\gamma)/d$

$$\lim_{n\to\infty} L_\gamma(X_n)/n^\alpha \quad = \quad \beta_{L_\gamma,d} \, \exp\left((1-\alpha)H_\alpha(f)\right), \qquad (a.s.)$$

Asymptotics of estimators of $H_\alpha(f)$

Define $B_p^{\sigma,q}$, the Besov space of $\ell_p(\mathbf{R}^d)$ functions with smoothness given by parameters σ and q.

Proposition 1 *Let $p > d \geq 2$ and $\alpha = (d-\gamma)/d \in [1/2, (d-1)/d]$*

$$\sup_{f^\alpha \in B_p^{1,1}} E^{1/\kappa}\left[\left|\int \widehat{f}^\alpha(x)dx - \int f^\alpha(x)dx\right|^\kappa\right] \;\geq\; O\left(n^{-1/(2+d)}\right)$$

while,

$$\sup_{f^\alpha \in B_p^{1,1}} E^{1/\kappa}\left[\left|\frac{L_\gamma(X_1,\ldots,X_n)}{n^\alpha} - \beta_{L_\gamma,d}\int f^\alpha(x)dx\right|^\kappa\right] \leq O\left(n^{-\frac{\alpha\lambda(p)}{1+\alpha\lambda(p)}\frac{1}{d}}\right)$$

where $\lambda(p) = d+1-d/p$.

Note: minimal-graph estimator converges faster for all $\alpha \geq 1/2$

Application of MST to Image Registration

1. Extract features from reference and transformed target images:

$$X_m = \{X_i\}_{i=1}^m \quad \mathit{and} \quad Y_n = \{Y_i\}_{i=1}^n$$

2. Construct following MST function on X_m and Y_n

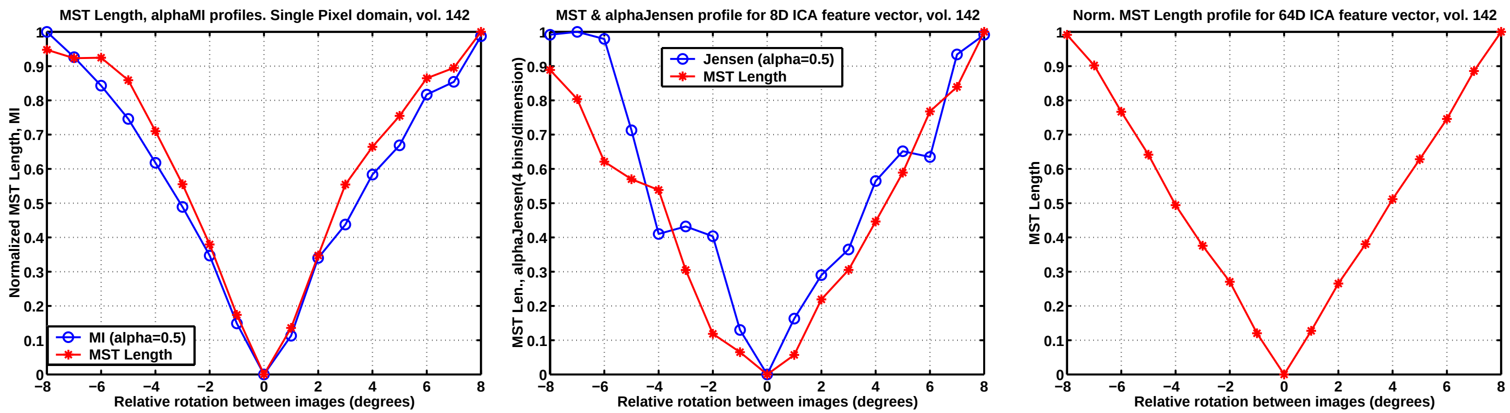
$$\Delta L = \ln L_\gamma(X_m \cup Y_n)/(n+m)^\alpha - \frac{m}{n+m}\ln L_\gamma(X_m)/m^\alpha - \frac{n}{n+m}\ln L_\gamma(Y_n)/n^\alpha$$

3. Minimize ΔL_γ over transformations producing Y_n.

$$(1-\alpha)^{-1}\Delta L \;\rightarrow\; H_\alpha(\varepsilon f_x + (1-\varepsilon) f_y) - \varepsilon H_\alpha(f_x) - (1-\varepsilon)H_\alpha(f_y)$$

where $\varepsilon = \frac{m}{m+n}$

Illustration for US Image Resgistration



(a) (b) (c)

Figure 9: Objective function profiles for histogram (L,M) and MST (L,M,R) estimators of α-Jensen difference vs histogram plug-in estimator ($\alpha = 1/2$): Single-pixel (L), 8D ICA (M), 64D ICA (R).

Quantitative Performance Comparisons

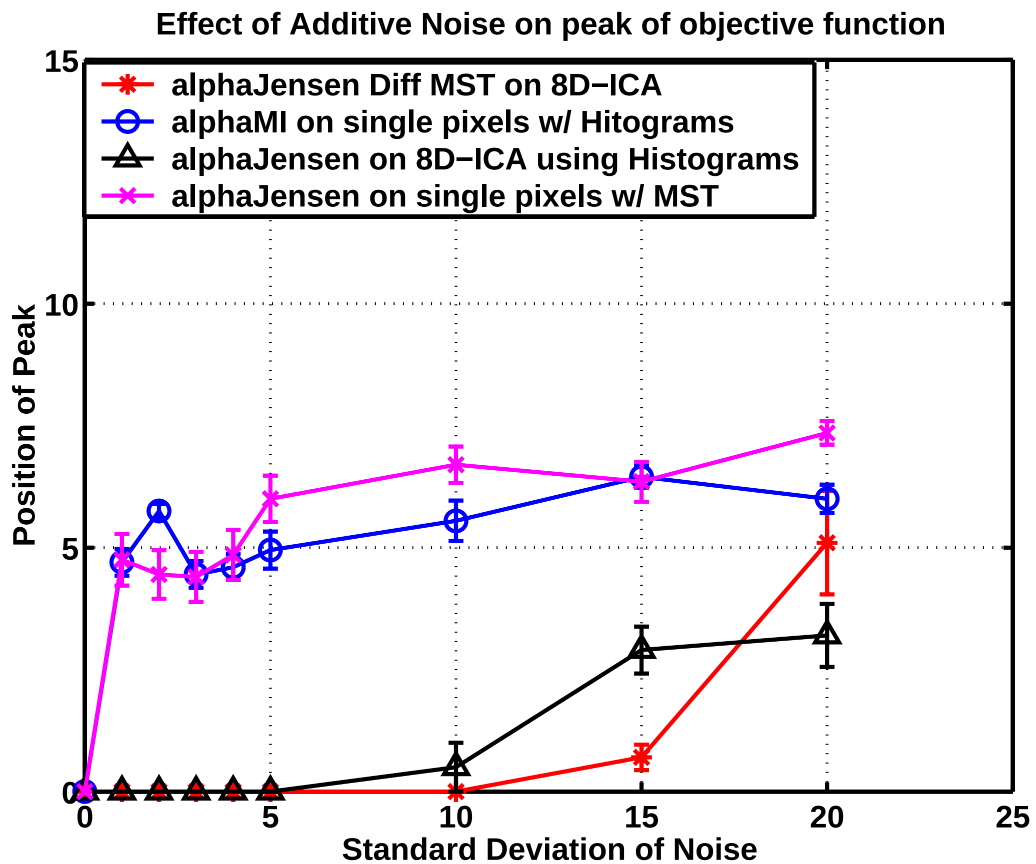


Figure 10: *Quantitative registration MSE comparisons.*

Extension: divergence estimation

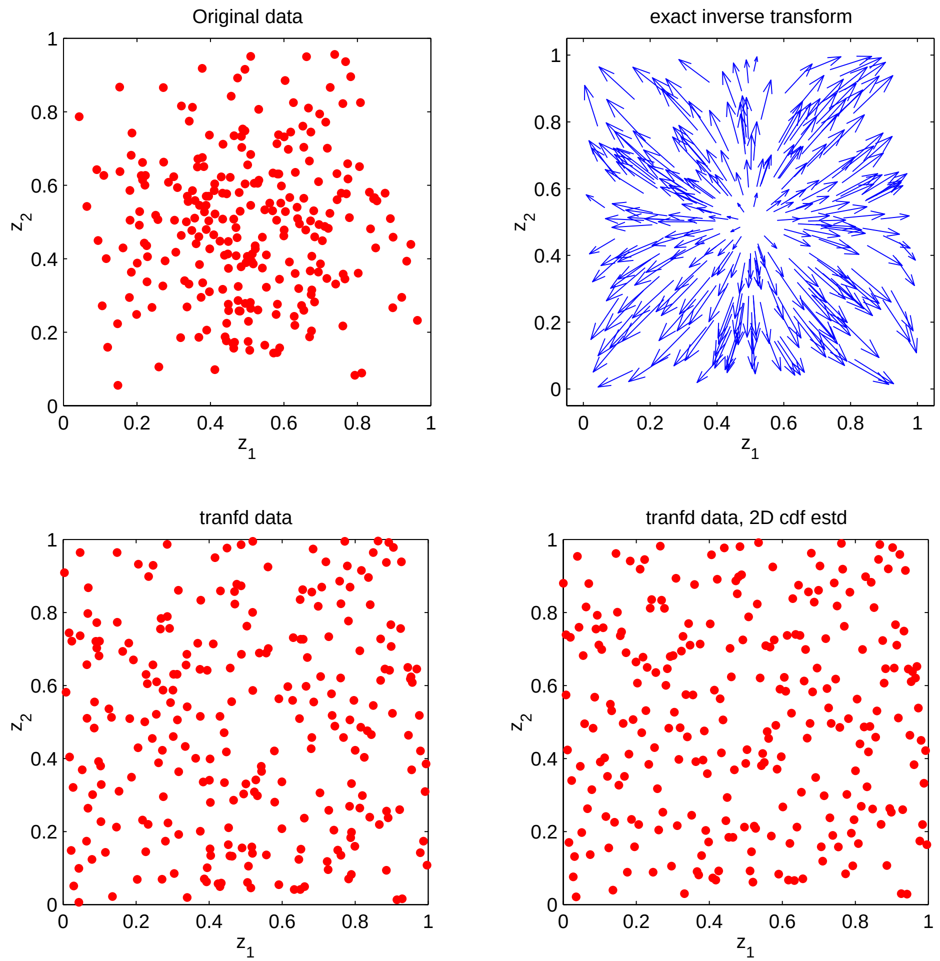


Figure 11:

Extension: outlier resistance via optimal pruning

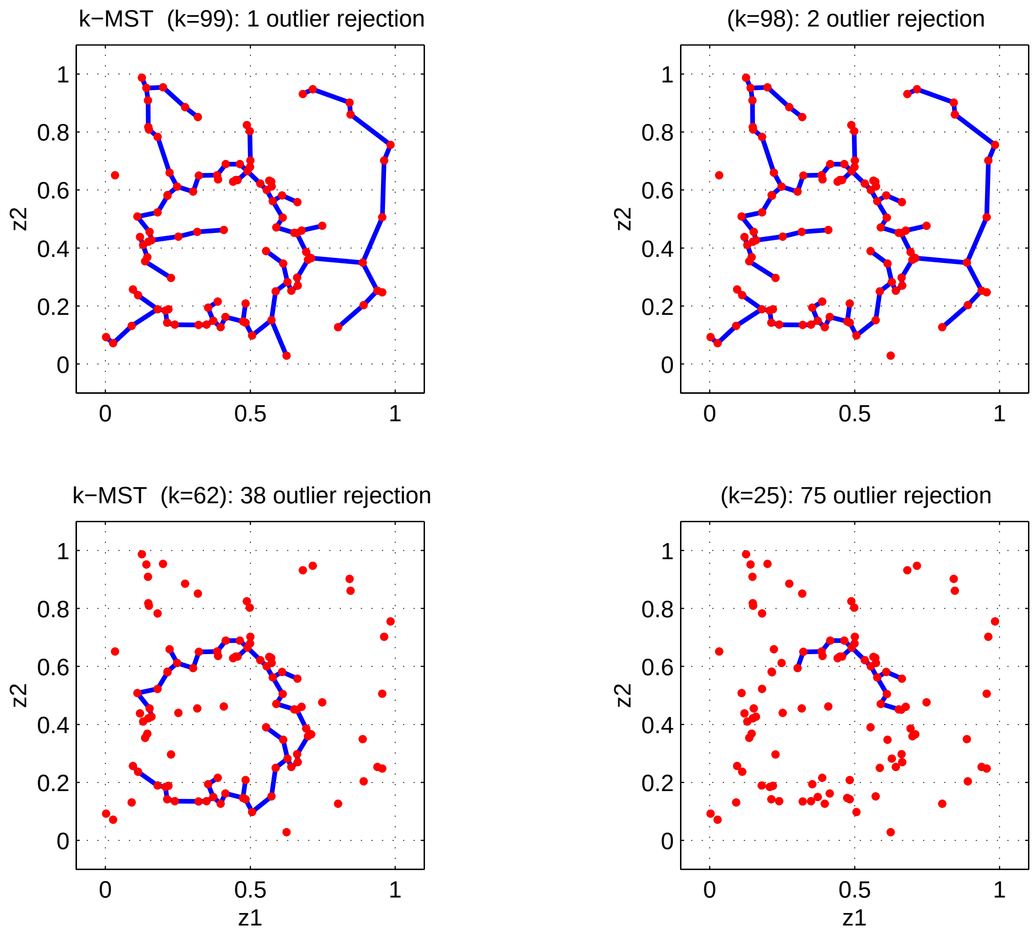


Figure 12: *k-MST for 2D annulus density.*

k-MST Influence Function for Gaussian Feature Density

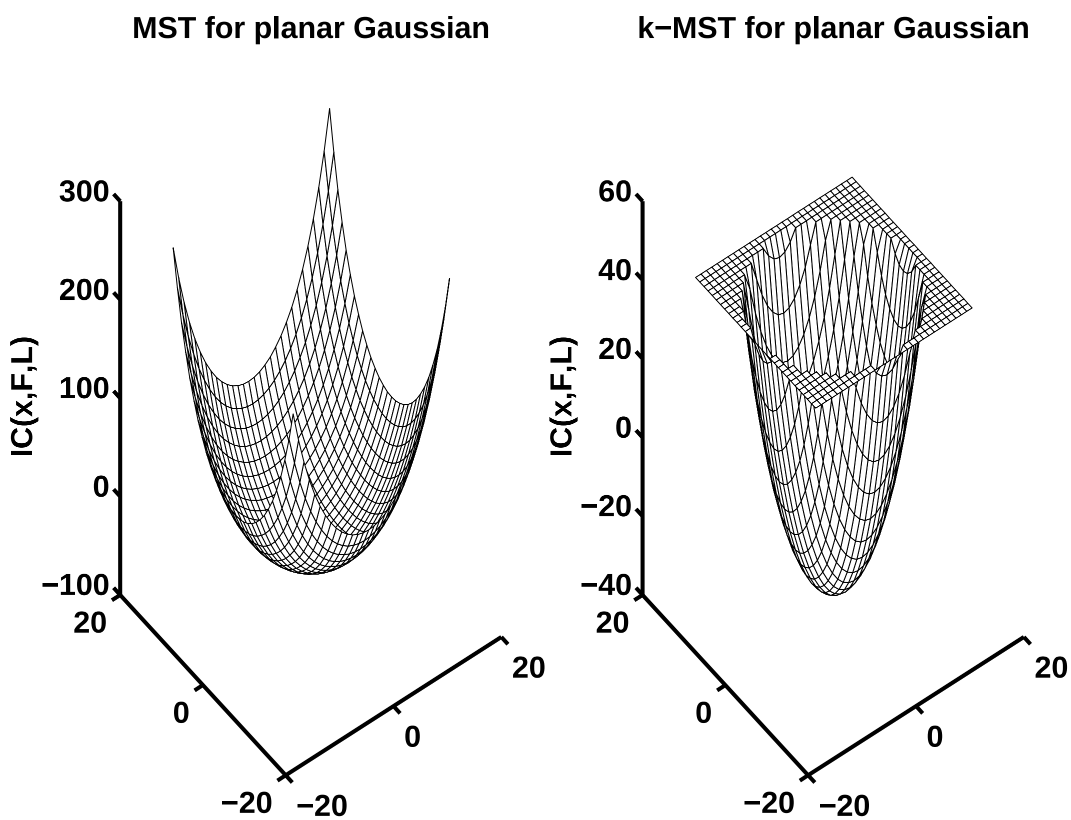


Figure 13: *MST and k-MST influence curves for Gaussian density on the plane.*

k-MST Stopping Rule

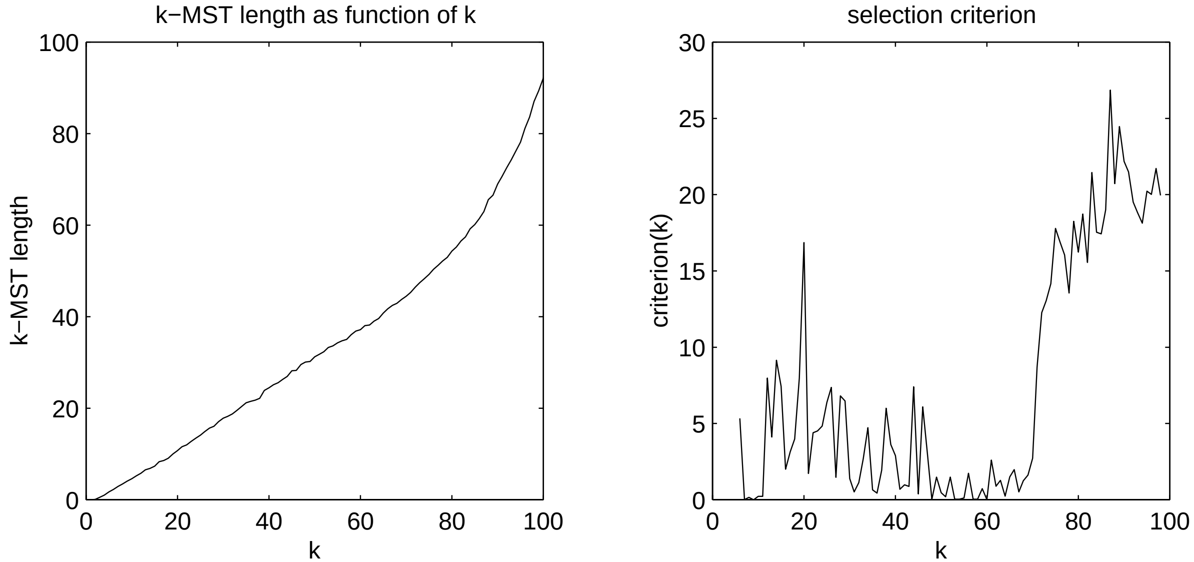


Figure 14: *Left: k-MST curve for 2D annulus density with addition of uniform “outliers” has a knee in the vicinity of* $n - k = 35$*. This knee can be detected using residual analysis from a linear regression line fitted to the left-most part of the curve. Right: error residual of linear regression line.*

Computational Acceleration of MST

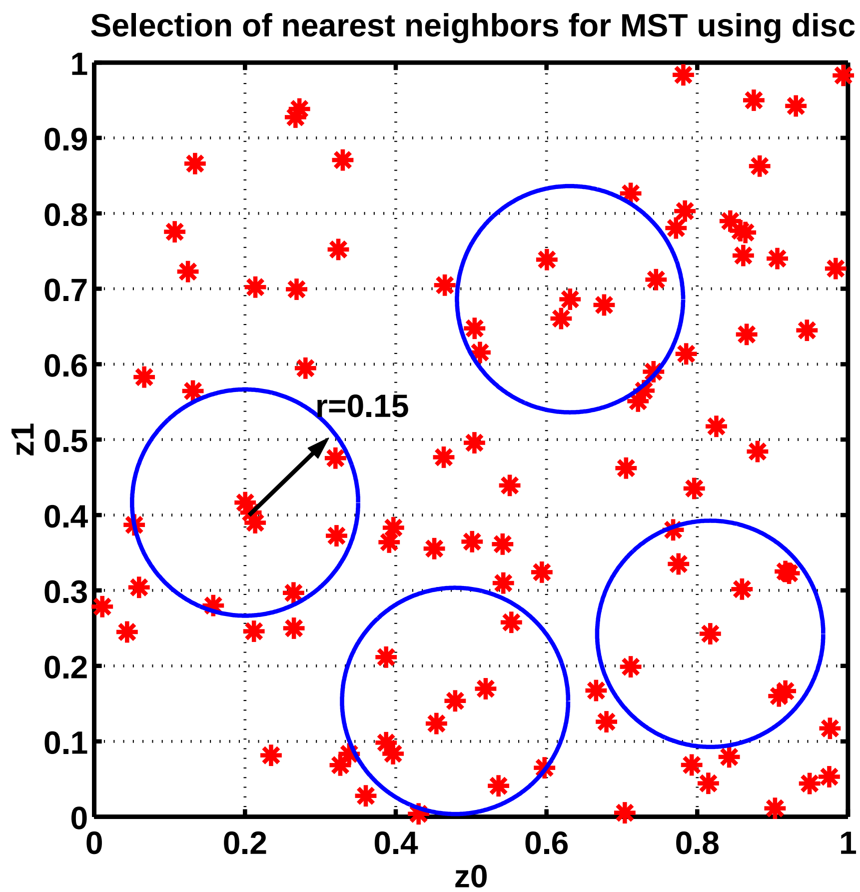


Figure 15: *Acceleration of Kruskal's MST algorithm from $n^2 \log n$ to $n \log n$.*

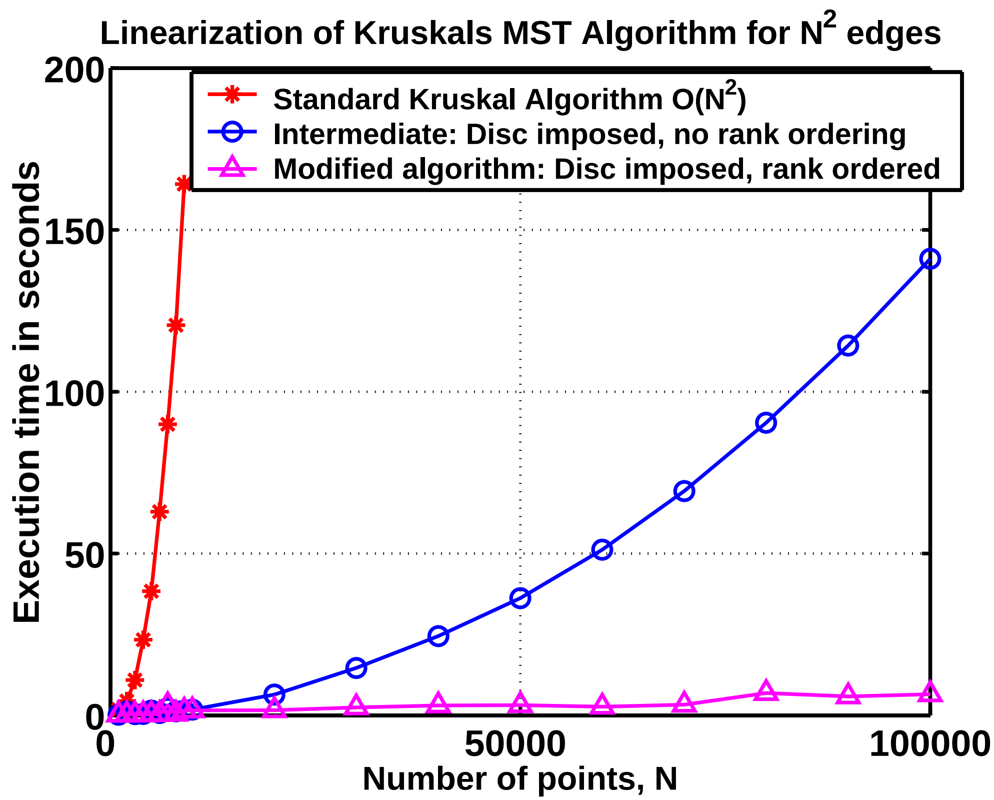


Figure 16: *Comparison of Kruskal's MST to our $n \log n$ MST algorithm.*

Conclusions

1. α-divergence for indexing can be justified via decision theory
2. Applicable to feature-based image registration
3. Non-parametric estimation is possible without density estimation via MST
4. MST outperforms plug-in estimation when latter is feasible
5. Robustified MST can be defined via optimal pruning of MST: k-MST