MICCAI 2010 Tutorial

Intensity-based Deformable Registration

Random Fields for Image Registration

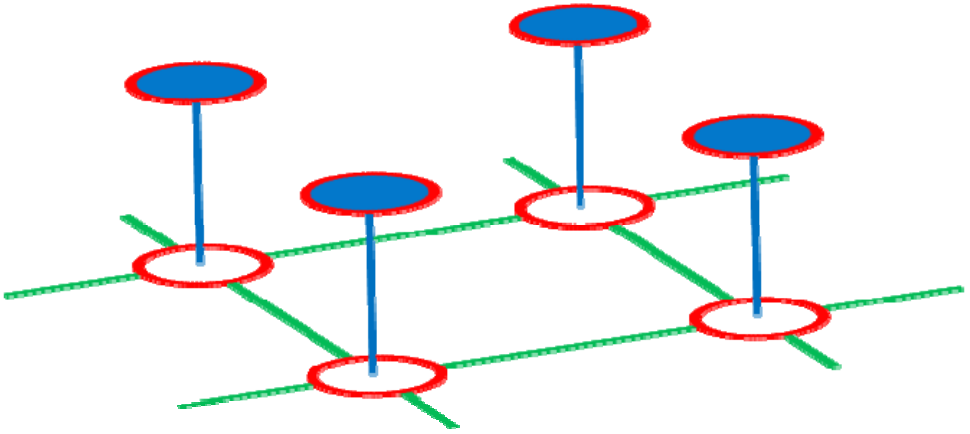
Bridging the Gap between
Continuous Transformations and Discrete Optimization

Ben Glocker

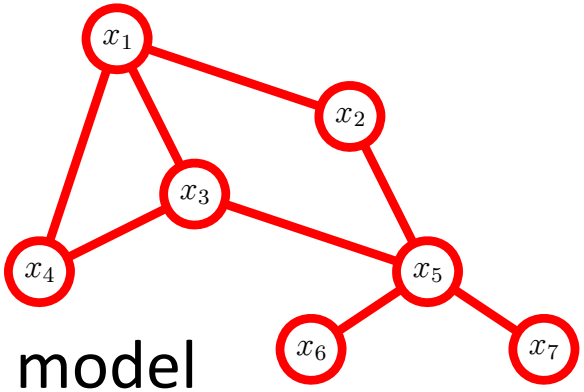
Computer Aided Medical Procedures (CAMP), TU München, Germany

PART I

Introduction to Random Fields



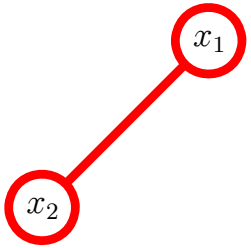
Random Fields



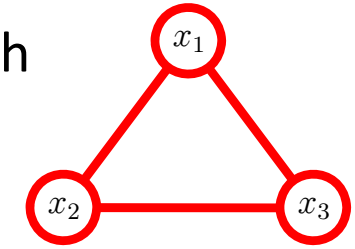
- A random field **X** is an undirected graphical model
  - corresponding graph $G=(V,E)$ with a set of nodes V and a set of edges E
- Each node $i \in V$ corresponds to a **random variable** X_i
- Random variables take values from a set of **labels** L_i
  - probability of a label assignment is $p(X_i=x_i)$ or $p(x_i)$ with $x_i \in L_i$
- Vector $\mathbf{x}=\{x_1,x_2,\ldots,x_n\}$ is called a **labeling** of the random field
  - the joint probability is denoted by $p(\mathbf{X}=\mathbf{x})$ or simply $p(\mathbf{x})$

Neighborhoods, Cliques, and Order

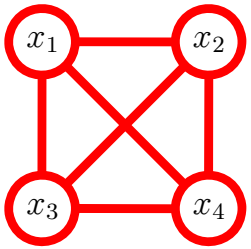
- Two nodes *i* and *j* are neighbors if $(i,j) \in E$



- A **clique** is a subset of nodes $C \subseteq V$
  - If $|C|>1$, every node $i \in C$ is a neighbor of all other $j \in C$
  - a clique is either a single node, or a fully-connected subgraph



- The **order** of a random field is $|C_{max}|-1$
  - first-order fields have cliques of size up to two
  - second-order fields have cliques of size up to three
  - etc…



Markov Random Fields

- Neighborhood system $N=\{ N_i \mid \forall\, i \in V \}$
  - N_i is the set of all neighbors of node i

Definition (MRF): A random field is said to be a Markov Random Field if and only if it satisfies the following two conditions

$$p(\mathbf{x}) > 0 \quad \forall\, \mathbf{x} \in \mathcal{X} \qquad \text{(Positivity)}$$

$$p(x_i | \{x_j : j \in N_i\}) = p(x_i | \{x_j : j \in V \setminus \{i\}\}) \qquad \text{(Markovian)}$$

Markov-Gibbs-Equivalence

- The Hammersley-Clifford-Theorem (1971) allows to define the joint probability as a factorization over the set of cliques **C**

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{C \in \mathbf{C}} \exp\left(-\psi_C(\mathbf{x})\right)$$

- The functions ψ_C are the so called **potential functions**
  - potential functions evaluate the quality of a clique labeling

Posterior, Likelihood, and Prior

- Often, some of the variables are **observed (y)** while others are unobserved or **hidden (x)**

- We are interested in the **posterior** probability $p(\mathbf{x}|\mathbf{y})$
  - posterior probability is hard to compute

- Rules of probability yield

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{x},\mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$$

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

Maximum A Posteriori

- The labeling with the maximum posterior probability

$$\begin{aligned}\hat{\mathbf{x}} &= \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) \\ &= \arg\max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \\ &= \arg\max_{\mathbf{x}} \frac{1}{Z} \prod_{C\in\mathbf{C}} \exp\left(-\psi_C(\mathbf{x},\mathbf{y})\right)\end{aligned}$$

- Problem: the computation of *Z* is often intractable

$$Z = \sum_{\mathbf{x}\in\mathcal{X}} \prod_{C\in\mathbf{C}} \exp\left(-\psi_C(\mathbf{x},\mathbf{y})\right)$$

Energy Minimization

- Convert the problem of maximum probability into an equivalent problem of minimum energy

$$\begin{aligned}\mathcal{E}(\mathbf{x}) &= -\log\left(p(\mathbf{x})\right) \\ &= -\log\left(\frac{1}{Z}\prod_{C\in\mathbf{C}}\exp\left(-\psi_C(\mathbf{x})\right)\right) \\ &= \sum_{C\in\mathbf{C}}\psi_C(\mathbf{x}) + \text{const}\end{aligned}$$

- The MAP labeling can be equivalently found by

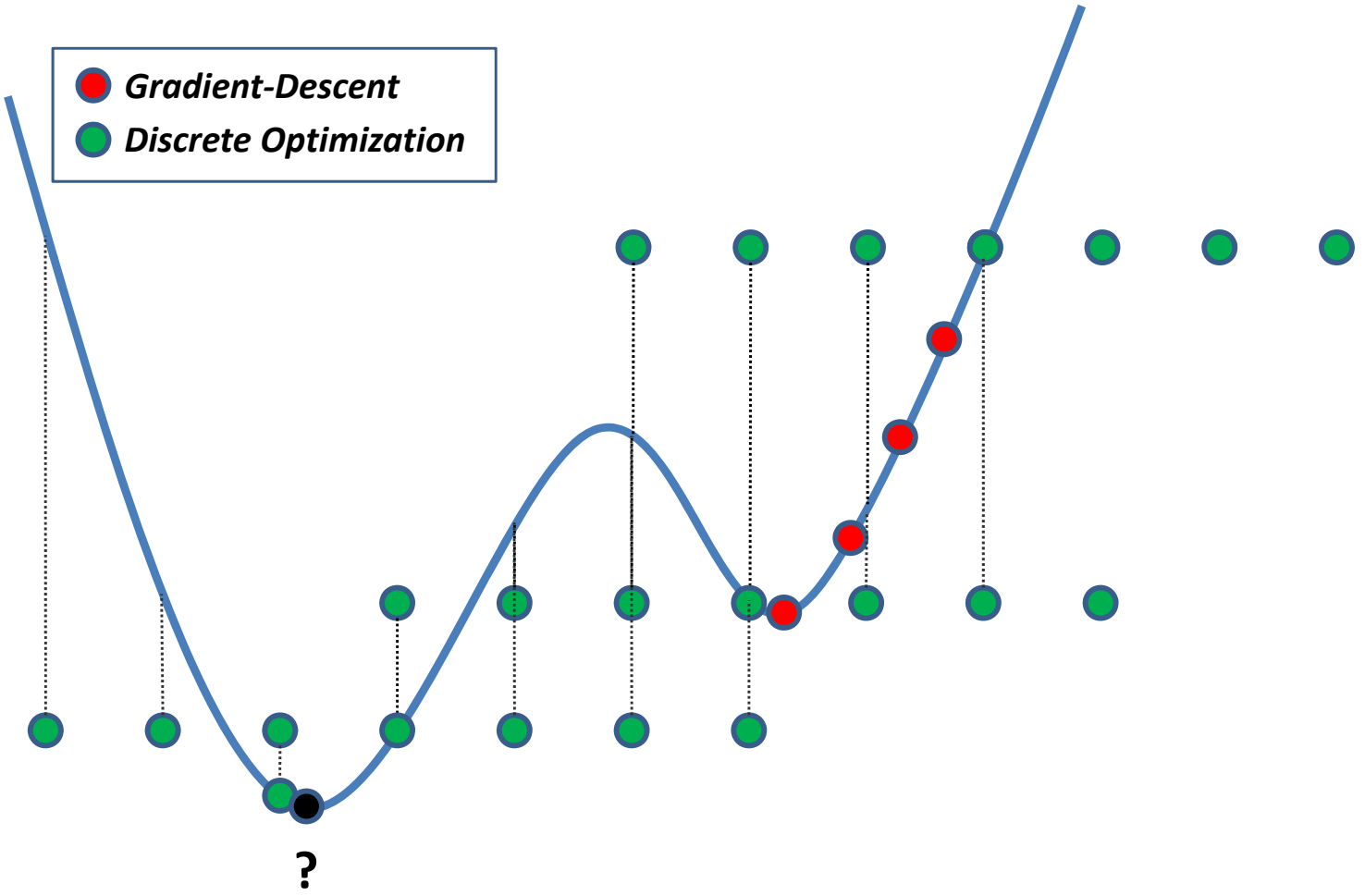
$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}}\mathcal{E}(\mathbf{x})$$

Discrete Markov Random Fields

- In discrete MRFs, the set of labels is finite
- Powerful combinatorial optimization techniques
  - Iterative graph-cuts: expansion moves, fusion moves, FastPD, …
  - Message-passing: efficient belief propagation, TRW-S, …
- Various applications in computer vision
  - segmentation, restoration, stereo, stitching, …

Motivation for Discrete Optimization

Optimization of Non-Convex Functions



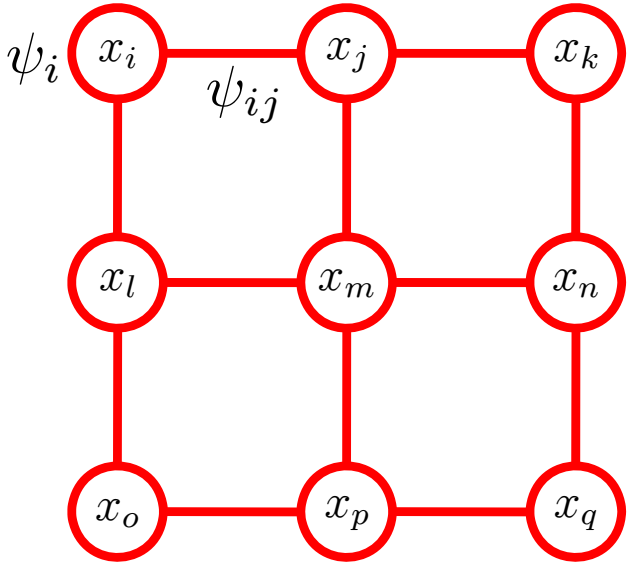
Example: Image Restoration

- Random variables correspond to pixels
- Labels correspond to intensities
- 4-connected pixel neighborhood

Observation *Result*

1st-order Random Field

$$E(\mathbf{x}) = \sum_{i \in C_1} \underbrace{\psi_i(x_i)}_{\text{unary pot.}} + \sum_{(i,j) \in C_2} \underbrace{\psi_{ij}(x_i, x_j)}_{\text{pairwise pot.}}$$



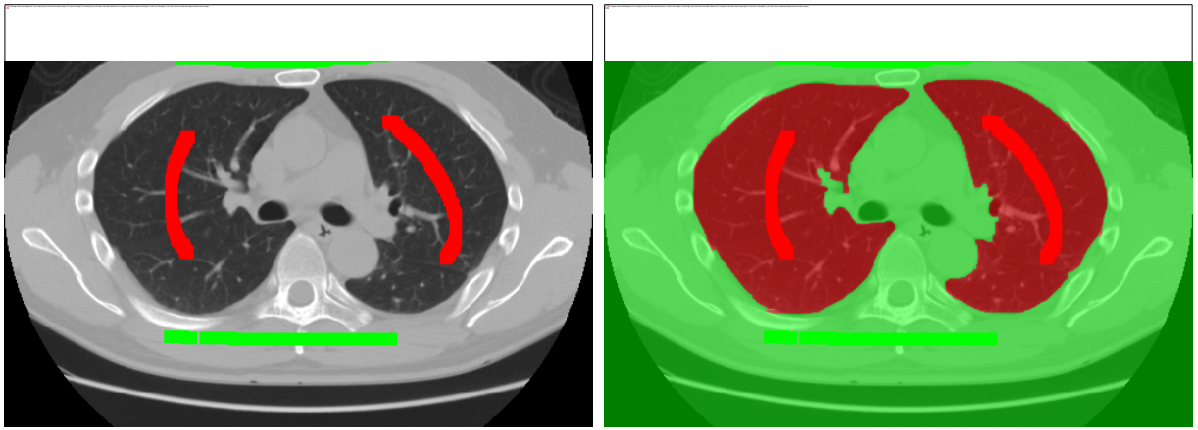
- Data Term: $\psi_i(x_i) = \begin{cases} 0 & \text{if } i \in \Omega_{\text{corrupt}} \\ (y_i - x_i)^2 & \text{otherwise} \end{cases}$
- Smoothness: $\psi_{ij}(x_i, x_j) = \lambda \min\left((x_i - x_j)^2, t\right)$

R. Szeliski, et al. ***A Comparative Study of Energy Minimization Methods for Markov Random Fields.*** *IEEE PAMI 2008*

Discrete Optimization in Imaging

Restoration

Stereo



Segmentation

R. Szeliski, et al. ***A Comparative Study of Energy Minimization Methods for Markov Random Fields.*** *IEEE PAMI 2008*
Y. Boykov and M.P. Jolly. ***Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images.*** *ICCV 2001*
S. Geman and D. Geman. ***Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images.*** *IEEE PAMI 1984*

Discrete Random Fields (in a nutshell)

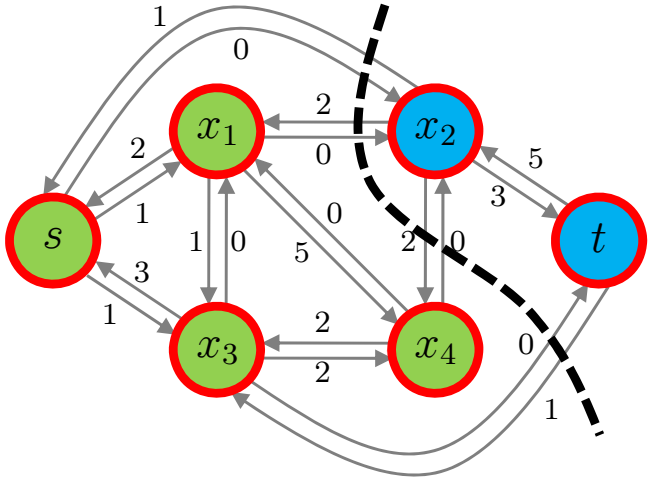
- Given ***n*** random variables $\mathbf{X}=\{X_1,\ldots,X_i,\ldots,X_n\}$
- a **clique** is a subset of variables $c \subseteq \mathbf{X}$
- the set of cliques C defines the **neighborhood system**
- a variable X_i can take a value x_i from a discrete set of **labels** L_i
- a **labeling x** is an assignment of values to all variables
- a **potential function** ψ_c evaluates the quality of a labeling of a clique
- the **energy** of a labeling is the sum of potential functions

$$E(\mathbf{x}) = \sum_{c \in C} \psi_c(\mathbf{x}_c) \qquad \hat{\mathbf{x}} = \arg\min_{\mathbf{x}} E(\mathbf{x})$$

S.Z. Li. ***Markov Random Field Modeling in Image Analysis.*** *Springer-Verlag 2009*
C.M. Bishop. ***Pattern Recognition and Machine Learning.*** *Springer-Verlag 2006*
R. Kindermann and J.L. Snell. ***Markov Random Fields and their Applications.*** *American Mathematical Society 1980*

PART II

Optimization via Iterative Graph-Cuts

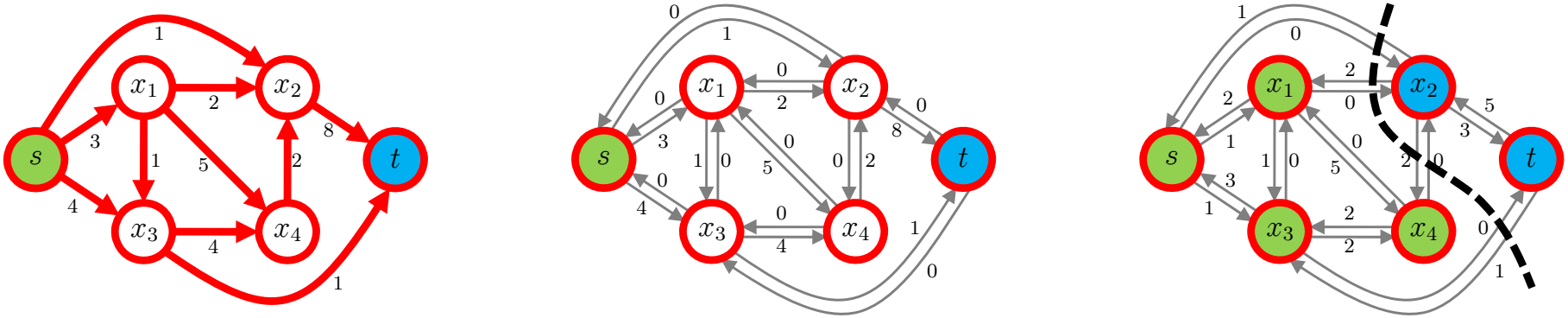


Graph-Cuts

- Graph-cut solves binary labeling problem, i.e. L={0,1}
  - applicable when energy is submodular
  - globally optimal via min-cut/max-flow algorithm

$$E(\mathbf{x}) = \sum_i \psi(x_i) \sum_{ij} \psi_{ij}(x_i, x_j)$$

$$\psi_{ij}(0,0) + \psi_{ij}(1,1) \leq \psi_{ij}(0,1) + \psi_{ij}(1,0)$$



V. Kolmogorov and R. Zabih. ***What energy functions can be minimized via graph-cuts?*** *IEEE PAMI 2004*

QPBO Graph-Cuts

- QPBO solves quasi-binary labeling problem, i.e. L={0,1,$\emptyset$}
  - quadratic pseudo-boolean optimization
  - applicable when energy is non-submodular
  - labels either with 0 or 1, or unlabeled = unknown

1. Persistency: *Let* ***y*** *be a binary labeling and let* ***z*** *be the "fusion" of the partial labeling* ***x*** *and* ***y*** *with* $z_i = x_i$ *if* $x_i \in \{0,1\}$ *and* $z_i = y_i$ *otherwise. Then,* $E(\mathbf{z}) \leq E(\mathbf{y})$.

2. Partial Optimality: *There exists a global minimum* ***y*** *such that* $x_i = y_i$ *for all labeled nodes in the partial labeling* ***x***

V. Kolmogorov and C. Rother. ***Minimizing non-submodular functions with graph-cuts – a review.*** *IEEE PAMI 2007*
C. Rother et al. ***Optimizing Binary MRFs via Extended Roof Duality.*** *CVPR 2007*

Expansion Moves

- Iterative graph-cuts solve multi-labeling, i.e. L={1,…,N}
- Each binary problem solves L={ "*keep*", "*change to alpha*"}
- Very large neighborhood search techniques

Algorithm 1: Alpha-Expansion

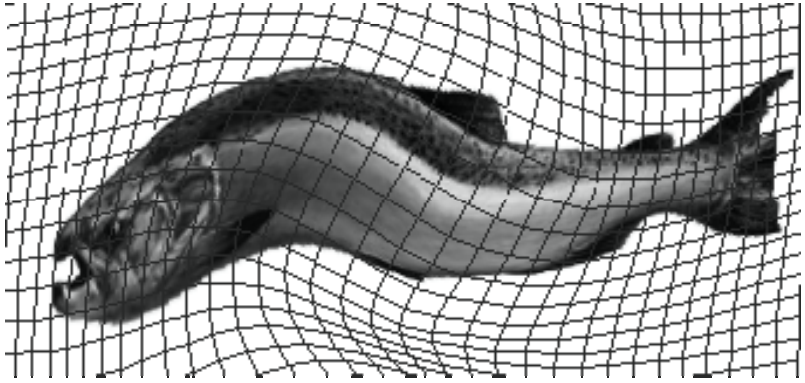
output: Labeling $\mathbf{x}$

```
1 x ← initializeLabeling() ;
2 for several sweeps do
3 |   foreach α ∈ L do
4 |   |   x ← x ⊙ α ;
5 |   end
6 end
```

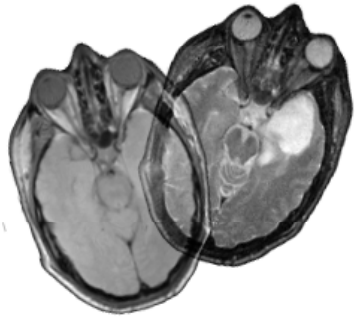
Boykov et al. ***Fast approximate energy minimization via graph cuts .*** *IEEE PAMI 2001*
N. Komodakis et al. ***Fast, Approximately Optimal Solutions for Single and Dynamic MRFs.*** *CVPR 2007*
Lempitsky et al. ***Fusion Moves for Markov Random Field Optimization.*** *IEEE PAMI 2010*

PART III

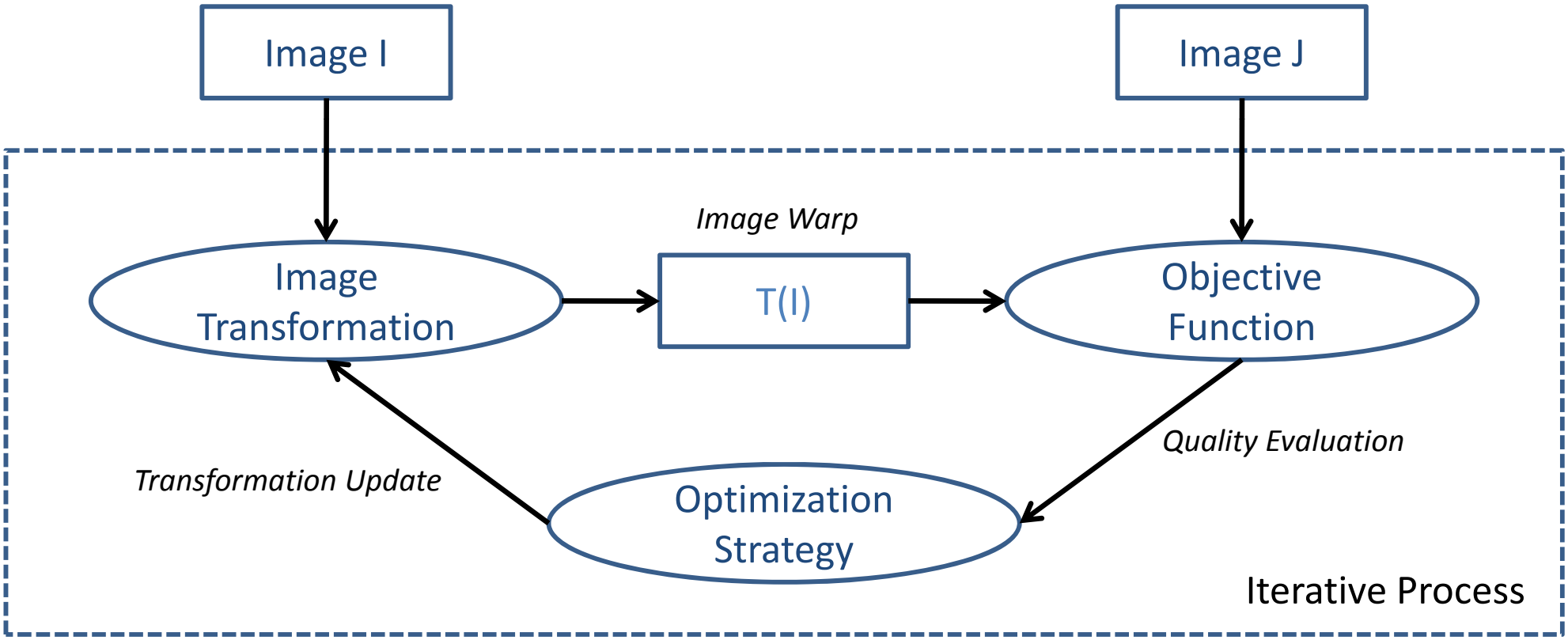
Non-Rigid Registration with Random Fields



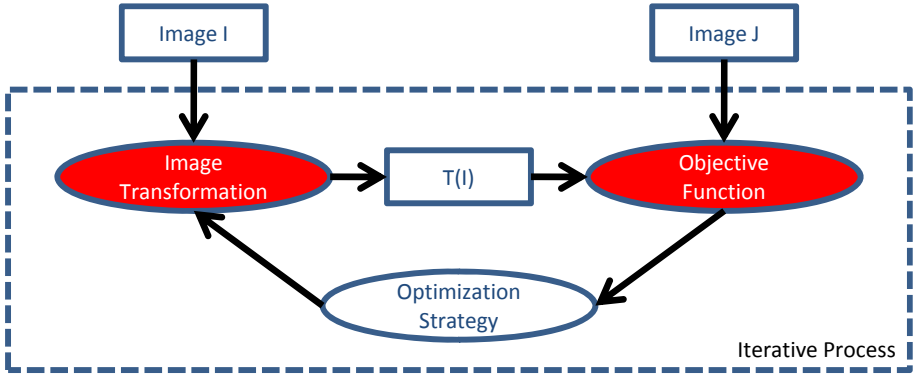
How do we do Registration?



Intensity-based Registration

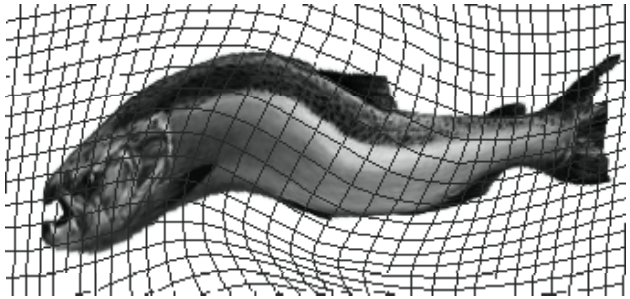


Non-Linear Case



Deformable Registration

$$T(\mathbf{x}) = \mathbf{x} + D(\mathbf{x})$$

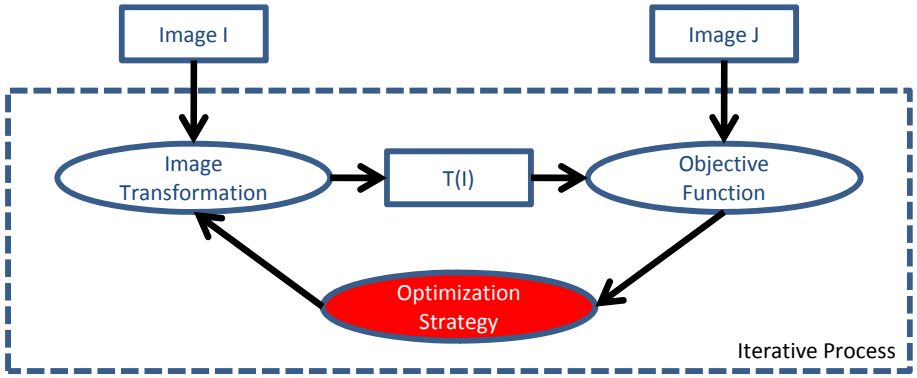


Ill-posed Problem: Need for Regularization Prior

$$E(I \circ T, J) = \underbrace{S(I \circ T, J)}_{\text{Similarity Measure}} + \underbrace{R(T)}_{\text{Smoothness Prior}}$$

A. Tikhonov. ***Ill-posed Problems in Natural Science.*** *Coronet. 1972*
J. Modersitzki. ***Numerical Methods for Image Registration.*** *Oxford University Press 2003*

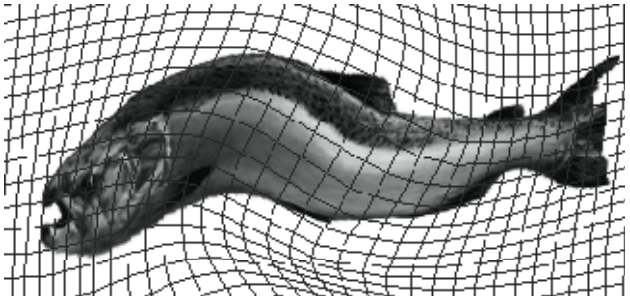
Optimization



Continuous Optimization	Discrete Optimization
Gradient-descent (& variants) Variational/Differential Approaches	Iterative Graph-Cuts Message-Passing
+ High-accuracy - Need differentiation of E - Hard to control - May converge to bad local minima	+ (Often) strong minima + Full control on search space - Accuracy bounded to discretization - Complexity for large problems

G. Hermosillo, et al. ***Variational Methods for Multimodal Image Matching.*** *Int'l Journal of Computer Vision 2002*
J. Modersitzki. ***Numerical Methods for Image Registration.*** *Oxford University Press 2003*
S. Klein, et al. ***Evaluation of optimization methods for nonrigid medical image registration using MI and B-splines.*** *IEEE Imag.Proc. 2007*
R. Szeliski, et al. ***A Comparative Study of Energy Minimization Methods for Markov Random Fields.*** *IEEE PAMI 2008*

Random Fields for Registration



- Random variables correspond to pixels
- Labels correspond to displacements

Curse of Dimensionality

Example: Non-linear Volume Registration

- e.g. 256^3 voxels → **16.7 millions of unknowns**
- search space is $\mathcal{R}^3$

Naive approach is inefficient and memory demanding!

T.W.H. Tang and A.C.S. Chung. ***Non-Rigid Image Registration using Graph-Cuts.*** *MICCAI 2007*
S. Roy and V. Govindu. ***MRF Solutions for Probabilistic Optical Flow Formulations.*** *Int'l Conf. Pattern Recognition 2000*

Summary of our Approach

Dimensionality Reduction

- Parameterization of the displacement field through a set of control points

Iterative Refinement Strategy

- Small label sets for efficient optimization
- Iterative refinement for high accuracy

Fast Approximation Scheme

- Efficient computation of the registration objective function

B. Glocker et al. ***Continuous Deformations Meet Efficient Optimal Linear Programming.*** *IPMI 2007*
B. Glocker et al. ***Dense Image Registration through MRFs and Efficient Linear Programming.*** *Medical Image Analysis 2008*

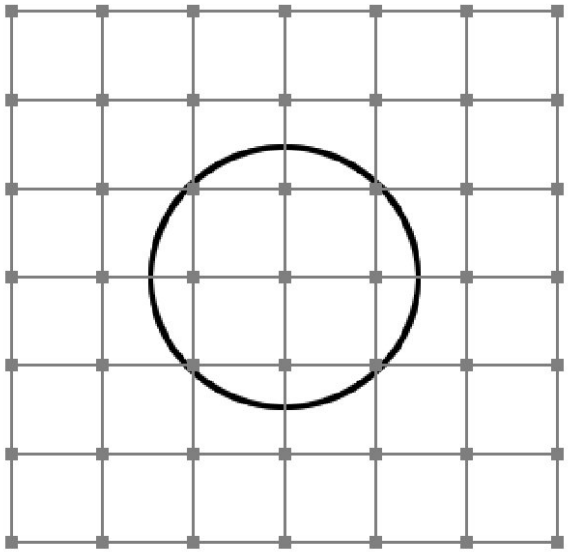
Dimensionality Reduction

- The dense displacement field can be represented by a sparse set of control points

$$T(\mathbf{x}) = \mathbf{x} + D(\mathbf{x})$$

$$D(\mathbf{x}) = \sum_i \eta_i(\mathbf{x})\,\mathbf{d}_i \quad \text{with} \quad \mathbf{d}_i \in \mathcal{R}^d$$

  - e.g. Free-Form Deformations using B-splines



Smooth Deformations and Coarse-to-Fine Hierarchy

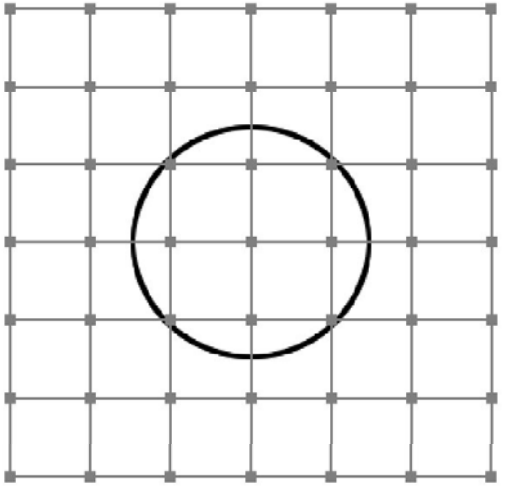
D. Rueckert et al. ***Nonrigid registration using free-form deformations: application to breast MR images.*** *IEEE Medical Imaging 1999*

Our Random Field Energy

Based on the dimensionality reduction, we can define the 1st-order energy for the control grid

- Random variables correspond to control points
- Labels correspond to control point displacements
- Registration prior on neighboring control points

$$E(\mathbf{x}) = \underbrace{\sum_{i \in C_1} \psi_i(x_i)}_{\text{Data Term: } S(I \circ T, J)} + \underbrace{\sum_{(i,j) \in C_2} \psi_{ij}(x_i, x_j)}_{\text{Smoothness: } R(T)}$$



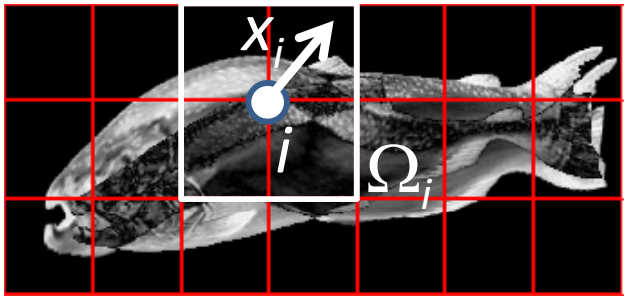
Data Term

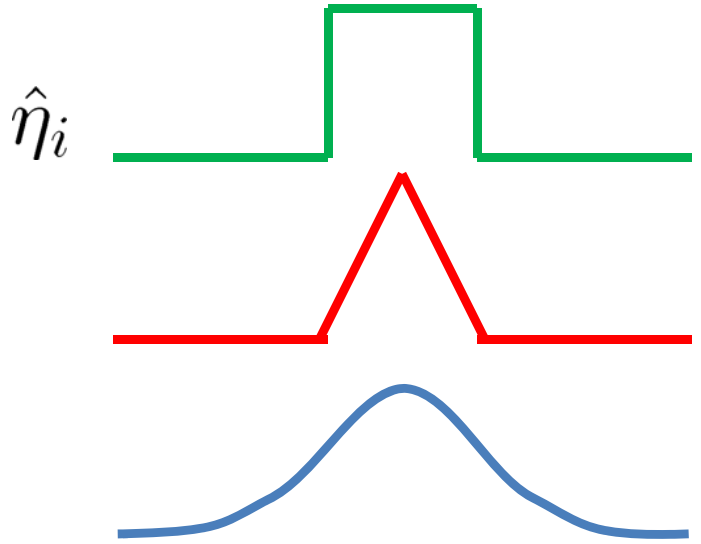
- Encodes local intensity-based similarity measure

$$\psi_i(x_i) = S_i(I \circ T_{x_i}, J)$$

  - Exemplary: mean squared intensity differences

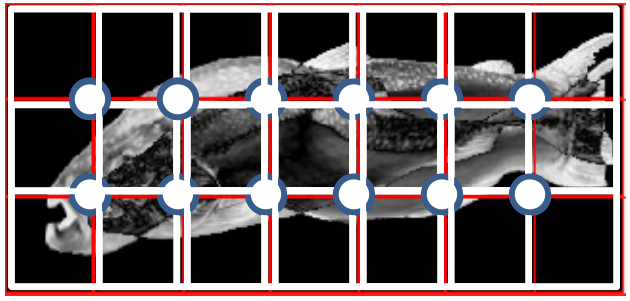
$$\psi_i(x_i) = \frac{1}{|\Omega_i|} \sum_{\mathbf{p} \in \Omega_i} \hat{\eta}_i(\mathbf{p}) \; (I(\mathbf{p} + D_{x_i}(\mathbf{p})) - J(\mathbf{p}))^2$$

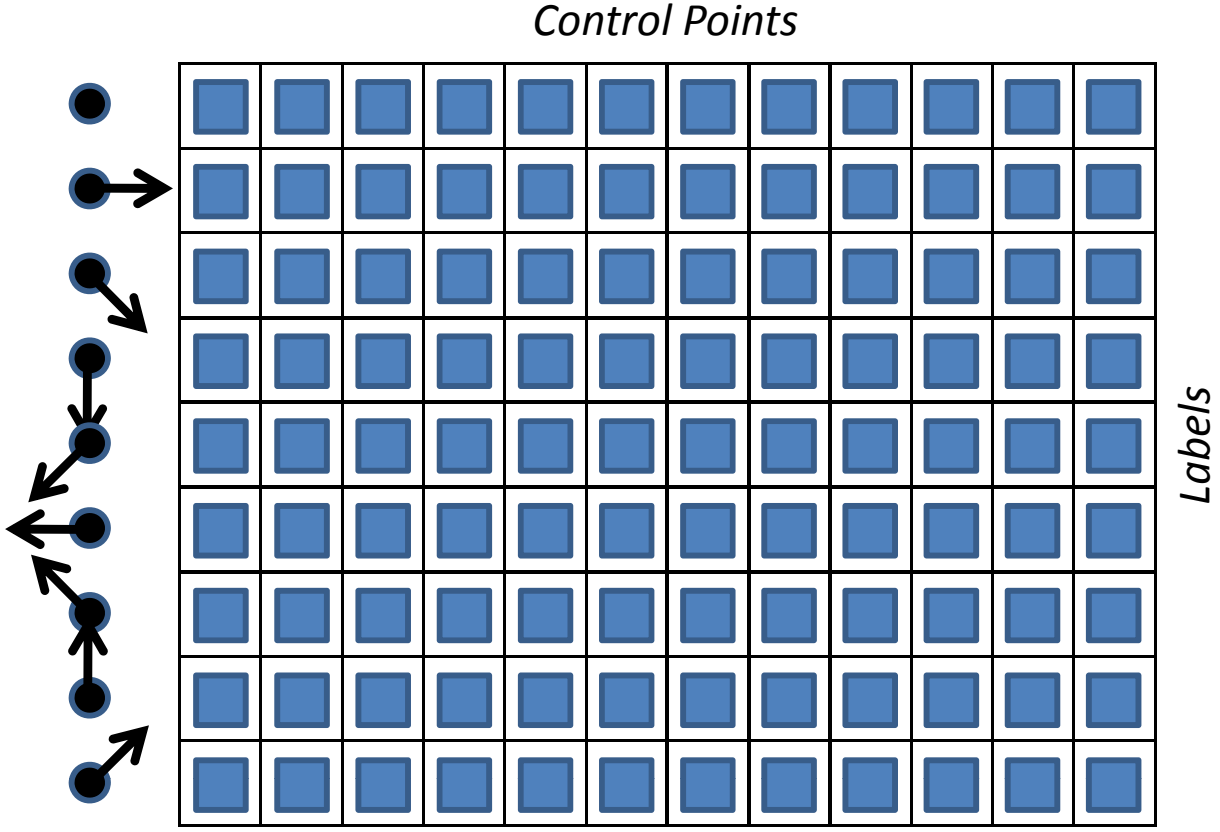




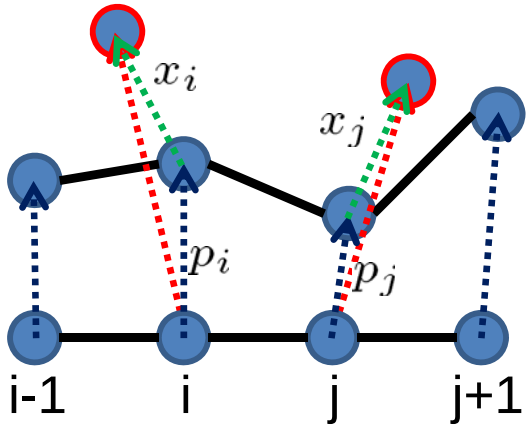
Fast Approximation Scheme

- Assume constant motion in local patch
- Compute the local similarity measures for one displacement/label simultaneously for all control points





Regularization Prior

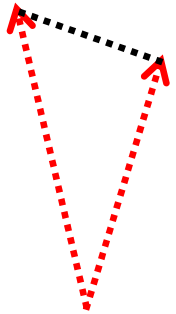


- Encodes local smoothness

$$\psi_{ij}(x_i, x_j) = R_{ij}(T_{x_i x_j})$$

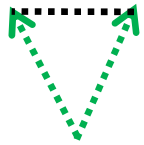
- (Truncated) absolute/quadratic difference: 1st derivatives
  - on the displacement field (***elastic-like***)

$$\psi_{ij}(x_i, x_j) = \lambda \min\left(\frac{\|(p_i + x_i) - (p_j + x_j)\|}{\|i - j\|}, t\right)$$



  - on the update field (***fluid-like***)

$$\psi_{ij}(x_i, x_j) = \lambda \min\left(\frac{\|x_i - x_j\|}{\|i - j\|}, t\right)$$



B. Glocker et al. ***Approximated Curvature Penalty in Non-rigid Registration using Pairwise MRFs.*** *Int'l Symp. Visual Computing 2009*
D. Kwon et al. ***Nonrigid Image Registration Using Dynamic Higher-Order MRF Model.*** *Europ. Conf. on Computer Vision 2008*

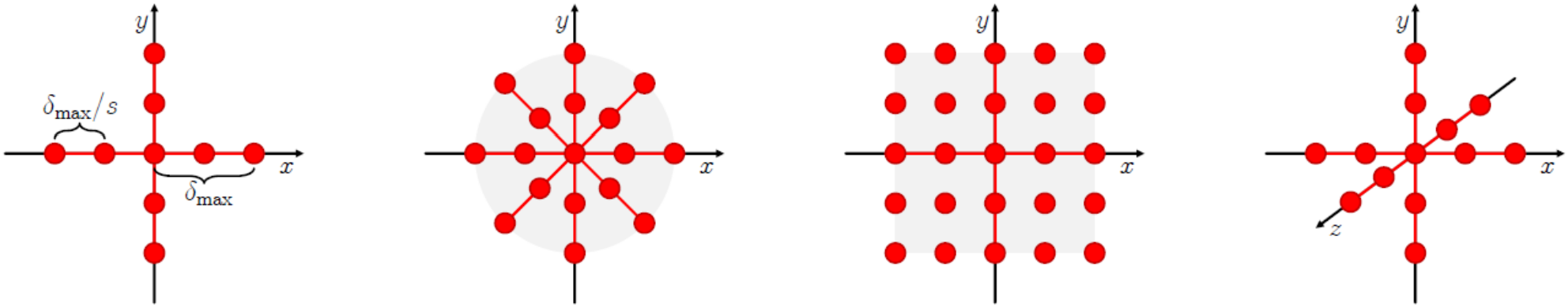
Search Space Discretization

- We need to discretize the continuous search space
  - definition of the **range** [-max;max]
  - the **number of samplings** s: |L|=2s+1
  - a **scaling factor** for the iterative refinement

Small Label Sets and Refinement for High Accuracy

+

Full Control on the Search Space



Our Basic Registration Algorithm

```
Input: Image I,J
Output: Transformation T
0: set initial T
1: set initial label space L
2: until convergence do repeat
3:     ΔT:= solve_discrete_labeling
4:     if( E(I∘(T+ΔT) ,J) < E(I∘T,J) )
5:           T:= T+ΔT
6:     else
7:           L:= refine_label_space
8:     end
9: end
```

*The actual implementation includes a **hierarchical approach** using **image pyramids** and **control grid refinement**!*

Live Demo

Breathing Motion Compensation

http://www.mrf-registration.net

Thank you! Questions?