Medical Image Registration using Information theory measures

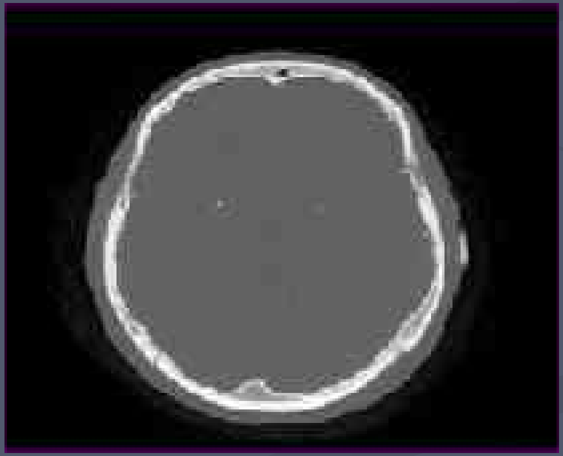
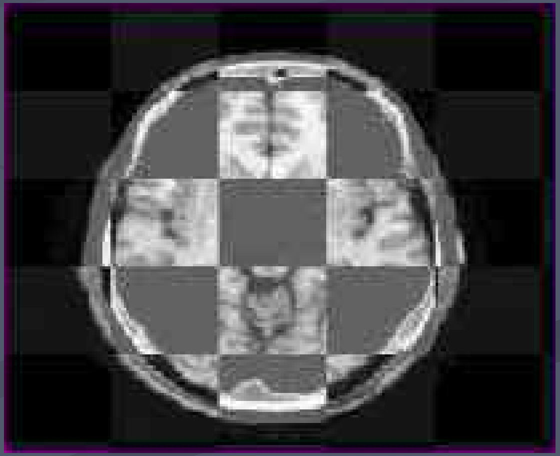
Ivan Vnučko
KAI
FMFI UK Bratislava
2005

Outline

- Introduction
- Classification
- Registration process
  - Transformation model
  - Similarity metric
  - Optimization method
- Challenges
- Recent methods

Image registration

- Process of estimating an optimal transformation between two images or image and a model
- Optimal transformation – optimum of some measure

Medical applications

- Fusion of data from different modalities
  - Anatomical (CT, MRI,…)
  - Functional (fMRI, SPECT, PET,…)
- Studying time changes
  - Monitoring progress of disease, comparison of pre- and post-intervention
- Image guided therapy
  - Planning surgery or radiotherapy on images registered to patient
- Atlas
  - Classification of structures based on registration to atlas

Classification

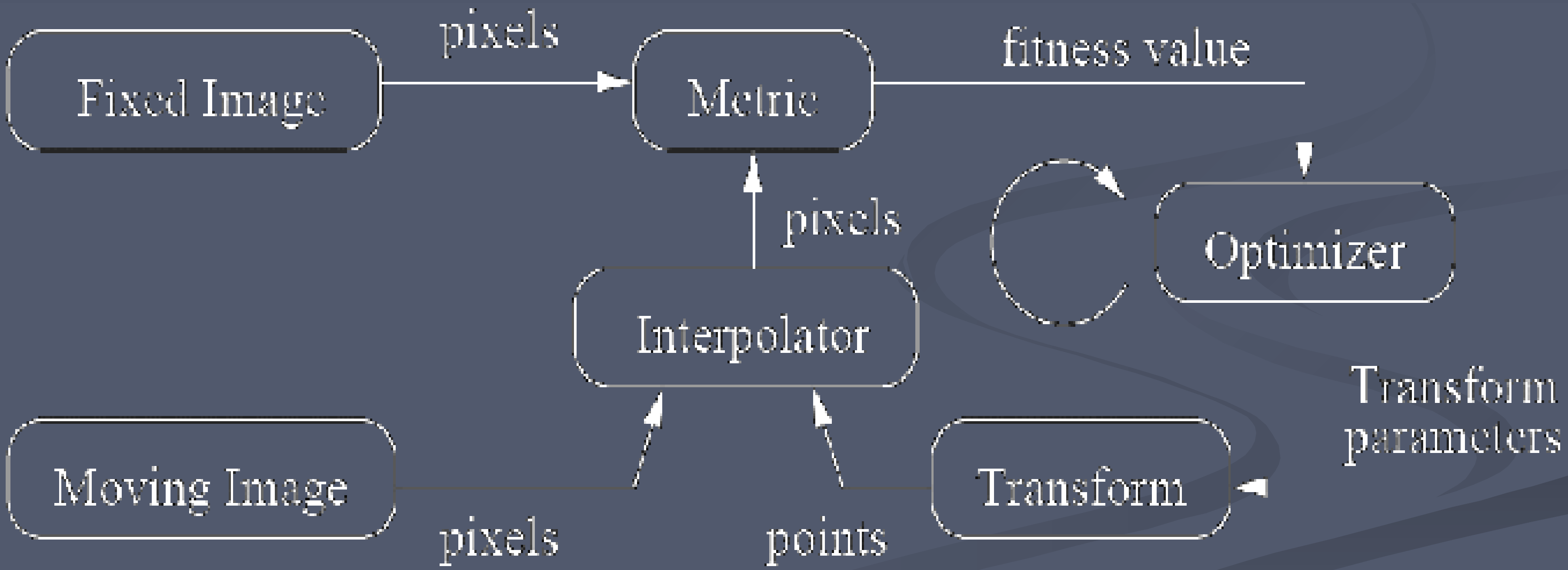
- 9 criteria classification Van den Elsen, Pol, Viergever
- Registration basis
  - Extrinsic
    - Invasive – fiducials, stereotactic frame
    - Non-invasive – adapters, markers
  - Intrinsic
    - Landmark – anatomical, geometrical
    - Segmentation – rigid, deformable
    - Voxel–property based – properties of the image content
  - Non-image based – calibrated sensors

Classification II.

- Other criteria:
  - Dimension – 2D/2D, 2D/3D, time
  - Nature of transformation – rigid, affine, projective, curved
  - Domain of transformation – global, local
  - Interaction – interactive, semi-automatic, automatic
  - Optimization procedure – parameters computed, searched for
  - Subject – intra-subject, inter-subject, atlas
  - Object
  - Type of registration – multimodality, monomodality,…

Registration procedure

- Processing flow:



Transformation model

- Transformation model
  - Rigid - 6 parameters (3D), simple
  - Affine – 12 parameters, crude approximation to non-rigid
  - Projective, perspective
  - Non-rigid, elastic, curved
    - Diffeomorphisms (e.g. fluid models), B-splines, thin-plate splines, FFD
    - Use of smoothness constraints, regularization term
  - Local – piecewise registration, global
- Rectification – correction of distortions with known models prior to registration
  - Can simplify expected transformation model
  - CT gantry tilt – affine without translation
  - Izotropic scaling
  - Projections to 2D – mostly curved, for example ”pincussion” or barrel distortion
  - MR – non-uniform gradient field due to scanner imperfections or inductions

Optimization

- Most voxel-property based methods use searching for parameters – iterative process
- With derivatives or without derivatives
- Stochastic
  - Simulated annealing
  - Time-consuming
- Deterministic
  - Powell, Levenberg-Marquardt, …
  - Faster, but problems with many local extrema
  - Multi-scale techniques
- Capture range – searching for optima, which is global only in some range, could be local only

Voxel similarity measures

- Computed from all voxels or from a subset
  - Regular grid
  - Random voxels
  - Segmented ROI
  - Preprocessing – blur to avoid alias artifacts
- Computed from intensities or properties
  - Gradients,…
- Computed only from overlapping voxels
  - Actual transformation dependent
- Grouping of intensity values
  - bins, classes

Voxel similarity measures II.

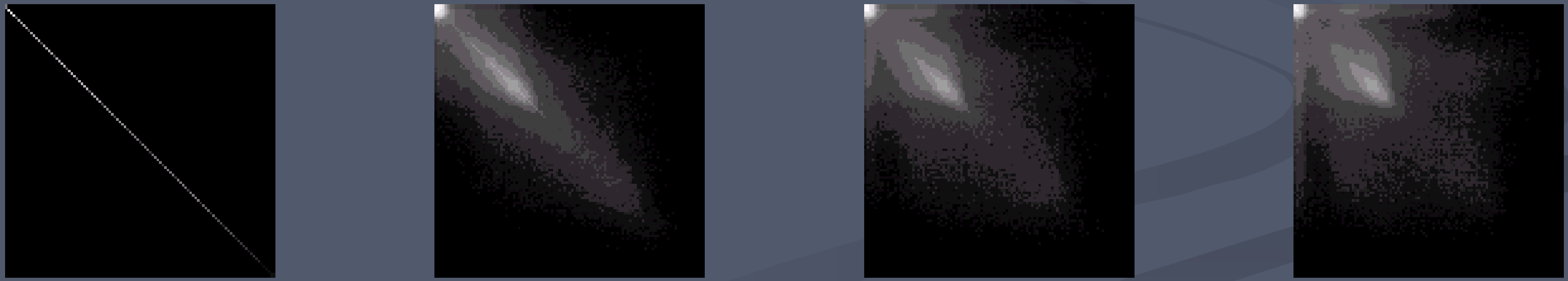
- Squares of intensity differences
  - Basic measure
  - Assumption: corresponding voxels have same intesities, for Gaussian noise it's optimum
  - Pre-segmentation – compute only on regions with small changes
- Correlation coefficient
  - Assumption: intensities are linearly related
- Ratio-image uniformity
  - Ratio-image – dividing all voxels in A by voxels in B
  - Minimizing normalized standard deviation
- Partitioned intensity uniformity (PIU)
  - Partitioned intensities in A, for each partition histogram of correspondent voxel intensities in B
  - Normalized standard deviation minimized

Joint histogram

- 2D histogram of co-occurrences of image values – joint histogram or feature space
- aligned images have focused histogram, by misalign dispersion grows

⇒ minimizing dispersion (some it's measure)

⇒ the best alignment



Joint probability distribution (pdf)

- Normalized joint histogram – joint pdf estimation
- Joint pdf changes with transformation
- Calculate measures from joint pdf to optimize transformation
- Used
  - Discrete estimation of joint pdf
    - normalized histogram
  - Continuous pdf estimation
    - Gaussian, Parzen windows

Entropy

- Measure of information, uncertainty
- Shannon – $H(X) = -\sum p_i \log p_i$
- Continuous RV – differential entropy
  $$H(X) = -\int p(x) \log(p(x))\, dx$$
- More than 20 other entropy definitions – functional
- Conditional entropy
  $$H(X) = -\int p(x) \log(p(x|y))\, dx$$
- Joint entropy
  $$H(X) = -\int p(x) \log(p(x,y))\, dx$$

Kullback – Leiber distance (KLD)

$$D(p,q) = \int p(x,y) \log \left(p(x,y)/q(x,y)\right) dxdy$$

- "distance" between pdf's, asymmetric
- KLD as an alignment measure:
  - minimize distance of data to some expected distribution
  - maximize distance of data to some unwanted distribution

Joint and conditional entropy

- maximizing the KLD of joint pdf from constant distribution is minimizing joint entropy:

$$D(p,c) = \int p(x,y) \log (p(x,y)/c)\, dxdy \approx -H(X,Y)$$

- maximizing the KLD of joint pdf from one of the marginal distributions is maximizing conditional entropy:

$$D(p,cm) = \int p(x,y) \log (p(x,y)/cm(x))\, dxdy \approx H(X|Y)$$

- maximizing the KLD of joint pdf from joint pdf if marginal were independent (m(x)n(y)) is maximizing mutual information:

$$D(p,mn) = \int p(x,y) \log (p(x,y)/m(x)n(y))\, dxdy = I(X,Y)$$

Mutual information

- Maximizing the distance of joint distribution from joint distribution of independent variables:

  $D(p,mn) = \int p(x,y) \log (p(x,y)/m(x)n(y))\, dxdy = I(X,Y)$

- $I(X,Y) = H(Y) - H(Y|X) = H(X) - H(X|Y)$
  - MI is the loss of uncertainty in Y (or X) when X (or Y) is known
- $I(X,Y) = H(X) + H(Y) - H(X,Y)$
  - Maximizing the MI is equivalent to maximizing marginal entropies and minimizing the joint entropy (last term)
  - Advantage in using mutual info over joint entropy - individual input's entropies are included

Properties of MI

- I(A,B) = I(B,A)
  - Symmetry
- I(A,A) = H(A)
  - MI of a message to itself is it's entropy
- I(A,B) <= H(A), I(A,B) <= H(B)
  - Info each message contains about the other cannot be greater than the info they themselves contain
- I(A,B) >= 0
  - Cannot increase uncertainty in A by knowing B
- If A, B are independent then I(A,B) = 0

Improvements to MI concept

- Normalized MI
  - $NMI(A,B) = (H(A) + H(B))/H(A,B)$
  - Tries to overcome the overlap problem of MI
- Entropy correlation coefficient
  - $ECC(A,B) = 2 - 2/NMI(A,B)$
- Connected region labeling
  - Regions in M => label image L
    - automatic/manual segmentation
    - anatomical features
  - $I(M,L,N) = H(M) + H(L) + H(N) - H(M,L,N)$
  - M, L are registered – information between N and M,L
    - $I(M,L,N) = H(M,L) + H(N) - H(M,L,N)$

Challenges

- From Pluim, Maintz, Viergever in IEEE Transactions on medical imaging 2003:
- Curved registration
  - Physically realistic deformations, not only regularization
- Registration of 3 or more images
  - How to find optimum for more transformations, and if it exists
- Inter-subject registration
- Some combinations of modalities
  - Ultrasound

Challenges II

- Intraoperative registration
  - Patient position verification in radiotherapy, correction for tissue deformation
  - Fast matching to poor quality image with deformations
- Including spatial information to entropy based measures
  - Shannon's entropy is based on assumption, that all voxels are uncorrelated
- 2D/3D registration, specially slice to volume
  - Need for Digitally Reconstructed Radiographs (DDR)

Recent methods

- Cumulative residual entropy – Vang, Vemuri, Rao, Chen in proc. of ICCV’03
  - Alternative entropy definition
- Joint class histograms - Chan, Chung, Yu et al. in proc. of CVPR’03
  - Reduction of intensity values and their mapping to classes
- Information metric - Zhang, Rangarajan in proceedings of CVPR’04
  - “Near” MI measure, but a pseudometric
  - Extendable to multiple images case
- MI of Regions – Russakoff et al. in proc. of ECCV’04
  - Spatial information

Cumulative residual entropy (CRE)

- alternative entropy definition
- uses cumulative distribution function (cdf) instead of pdf:

$$H(X) = - \int P(|X|>\lambda) \log(P(|X|>\lambda))\, d\lambda$$

- can be computed from samples
- conditional CRE:

$$H(X|Y) = - \int P(|X|>x|Y) \log(P(|X|>x|Y))\, dx$$

- cross – CRE (CCRE):

$$C(X,Y) = H(X) – E(H(X|Y))$$

Information metric

- $\rho(X,Y) = H(X|Y) + H(Y|X)$
- Pseudometric
  - ($\rho(X,Y) = 0$ if $X = f(Y)$)
- $\rho(X,Y) = H(X,Y) – MI(X,Y)$
- Can be easily extended to multiple image case:
  - $\rho(X_1,\dots,X_n) = \sum^n H(X_i|X_1,\dots,X_{i-1},X_{i+1},\dots,X_n)$
  - Because of computational complexity of $\rho(X,Y,Z,\dots)$ upper bound for ρ (a metric too) used

Joint class histogram

- Class labels - ψ_1, ψ_2
- Segmentation => voxel classifications
  - L_1: X_1 -> ψ_1, L_2: X_2 -> ψ_2
- Mapping relation – R: $\psi_1 \times \psi_2$
- Expected joint class histogram (EH)
  - for all x_2 from X_2 increase bin($L_1(x_1)$, $L_2(x_2)$)
  - x_1: randomly from X_1, so that bin($L_1(x_1)$, $L_2(x_2)$) is from R
  - normalize
- Observed class histogram (O(T)) for transformation T
- Measure - KLD(O(T), EH)

MI of Regions

- Adding spatial information to MI concept
- Vector of neighbors of a pixel
- Curse of dimensionality