Image warping for the low-rank representation of batches of images

Oleg Makhnin

New Mexico Tech

SAMSI LDHD Transition Workshop,
May 12, 2014

Goal:

* Develop a method for aligning batches of images using a non-affine warp (or "nonlinear domain transformation").

* RASL (Robust Batch Alignment of Images by Sparse and Low-Rank Decomposition): Peng et al (2012). Uses affine transformations.

* Leads to a low-dimensional representation of a batch (dimension reduction), ignoring the "non-essential" variability.

* Image registration.

Related work – Refs not exhaustive!

* Optical flows Horn & Schnuck (1981), Lucas & Kanade (1981), Brox (2004), ...

Construct a "velocity field" moving pixels of one image to match the pixels of the other image. Not necessarily a diffeomorphism of image domain.

* LDDMM Trouvé(1998), Joshi(2004), Beg(2005), ...
Construct a path in the manifold of diffeomorphisms from one image to the other. Can develop an "average" image of a batch, or construct a path corresponding to movement along geodesics (e.g. age progression).

* SIFT/SURF Lowe (2004), Bay (2008),...

Local descriptor methods: match local features, then organize them together.

RASL review

* What I like about RASL:

- Not a single "mean" or "reference" image, but a (low-dimensional) subspace.

- L_1 norm is used for error ("sparse"): more robust against occlusions, misaligned edges etc.

Goal: Given several images, align them to a common subspace.

Peng'12:

$$\min_{A,E,\tau} \ rank(A) + \lambda \|E\|_0 \qquad \text{s.t. } D \circ \tau = A + E$$

Transformed images: $D_i \circ \tau_i$, assume they align to produce a low-rank template A.
The errors E are assumed sparse.

Let $D = m \times n$ matrix of images (stacked as columns), $\tau =$ set of image transformations, $D \circ \tau$: aligned images, $A =$ common template

Relaxation: $rank(A)$ is relaxed to nuclear norm (sum of singular values) and $\|E\|_0$ is relaxed to $\|E\|_1$. But

$$\min_{A,E,\tau} \|A\|_* + \lambda \|E\|_1 \qquad \text{s.t. } D \circ \tau = A + E$$

is a non-convex problem: $D \circ \tau$ is nonlinear even though τ's are affine.

Solve iteratively: given τ,

$$\min_{A,E,\Delta\tau} \|A\|_* + \lambda\|E\|_1 \quad \text{s.t.} \quad D\circ\tau + \sum_{i=1}^{n} J_i \Delta\tau_i \varepsilon_i^T = A + E \quad (1)$$

is a convex problem, then set $\tau_{\mathtt{new}} = \tau + \Delta\tau$.

(1) is solved using Augmented Lagrange Multiplier (ALM) Method.

Examples: faces (many recognition methods need to align faces first), video processing (image stabilization, tracking objects etc.)

No guarantees, but seems to work well if the initial misalignment is not too large.

Details

Solution: iterative linearization. $D \circ (\tau + \Delta\tau) \approx D \circ \tau + \sum_{i=1}^{n} J_i \Delta\tau_i \varepsilon_i^T$, J_i is Jacobian of ith image w.r.t. transformation τ_i and $\{\varepsilon_i\}$ is the standard basis for $\mathbb{R}^n$.

Here, τ_i are assumed to belong to some group $\mathbb{G}$ described by p parameters (e.g. $\mathbb{G} = SE(2)$ (translations + rotations), $GL(3)$ etc), J_i is then $m \times p$, $\tau = \mathtt{Stack}(\tau_i, i = 1, ..., n)$.
Let $h = D \circ \tau + \sum_{i=1}^{n} J_i \Delta\tau_i \varepsilon_i^T - A - E$. Then

$$\mathcal{L}_\mu := \|A\|_* + \lambda\|E\|_1 + \langle Y, h(A, E, \Delta\tau)\rangle + \frac{\mu}{2}\|h(A, E, \Delta\tau)\|_F^2$$

where Y is Lagrange Multiplier matrix and μ_k is an increasing sequence.
Note: $\langle X, Y\rangle = tr(X^T Y)$ and $\|\cdot\|_F$ is Frobenius norm

Then iterate

$$(A_{k+1}, E_{k+1}, \Delta\tau_{k+1}) = \arg \min_{A,E,\Delta\tau} \mathcal{L}_{\mu_k}(A, E, \Delta\tau, Y_k)$$

$$Y_{k+1} = Y_k + \mu_k h(A_{k+1}, E_{k+1}, \Delta\tau_{k+1})$$

$\min_{A,E,\Delta\tau}$ is done by alternating minimization w.r.t. A, E or $\Delta\tau$, using SVD with soft thresholding (shrinkage).
Three levels of iteration in all. Parameter λ is set to $1/\sqrt{m}$.

Theoretical convergence proven only for two unknowns in the alt. minimization scheme!

A MATLAB implementation is available, takes about 3 minutes on a 2.8 GHz Macbook Pro for 100 images, each 80×60.

penalized RASL

Transformations are parameterized $\tau(\xi)$, we need regularity

$$\min_{A,E,\tau} \|A\|_* + \lambda\|E\|_1 + \kappa C(\xi) \qquad \text{s.t. } D \circ \tau(\xi) = A + E$$

$C(\cdot)$ is a convex penalty term.

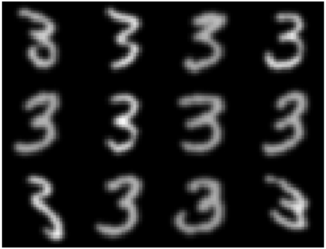
Then linearize: given ξ,

$$\min_{A,E,\Delta\xi} \|A\|_* + \lambda\|E\|_1 + \kappa J_C(\xi)\Delta\xi \qquad \text{s.t.} \qquad D \circ \tau + \sum_{i=1}^{n} J_i \Delta\tau_i(\Delta\xi)\varepsilon_i^T = A + E \tag{2}$$

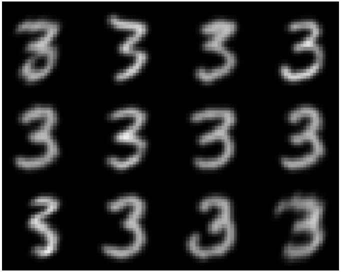
should work at least for small κ

Raw

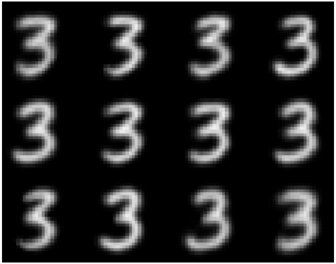
RASL



RASL-penalized



Low-rank A

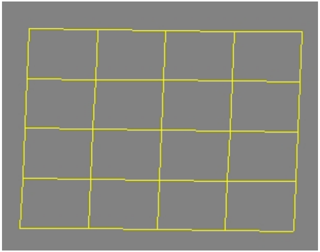


Example:

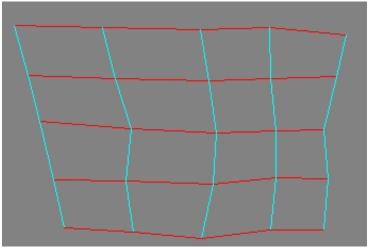
Handwritten digits (MNIST)

RASL-n (non-affine)

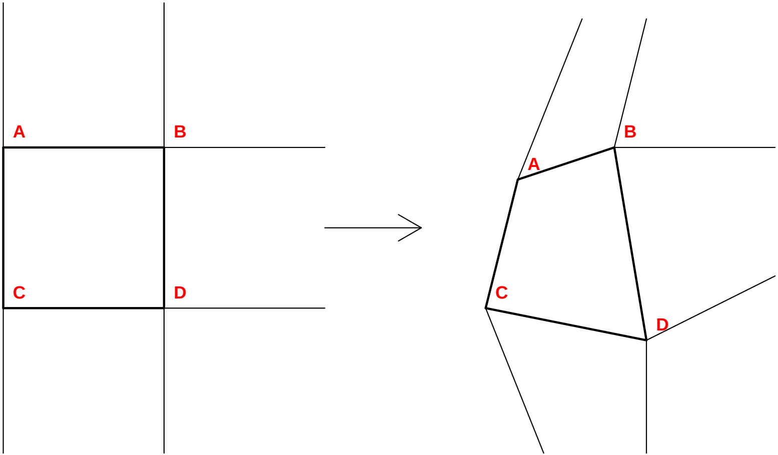
Extend to a piecewise-affine transform, with parameterization $\xi = x$ and y coords of new nodes into which the initial rectangular grid is transformed.



$\Longrightarrow$



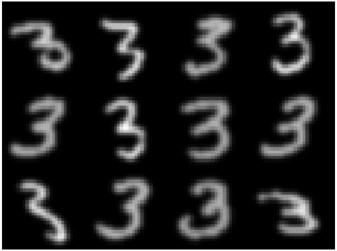
*Multiscale: start with an affine (RASL) alignment, then subdivide into 2x2 grid, then subdivide into 4x4 etc.



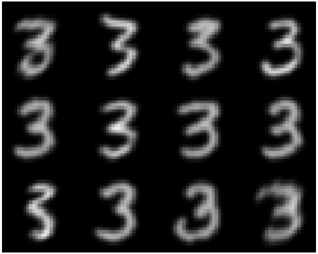
Penalty: a) orthogonality $(\overrightarrow{AB}\cdot\overrightarrow{AC})^2+(\overrightarrow{AB}\cdot\overrightarrow{BD})^2+\ldots$

b) length $((|AB|-\ell)_+)^2+((|AC|-\ell)_+)^2+\ldots$

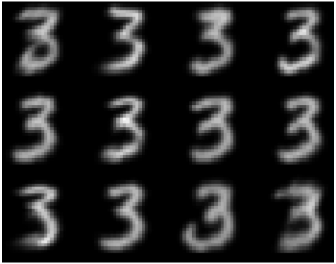
Raw



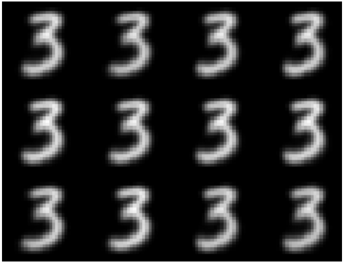
RASL-penalized

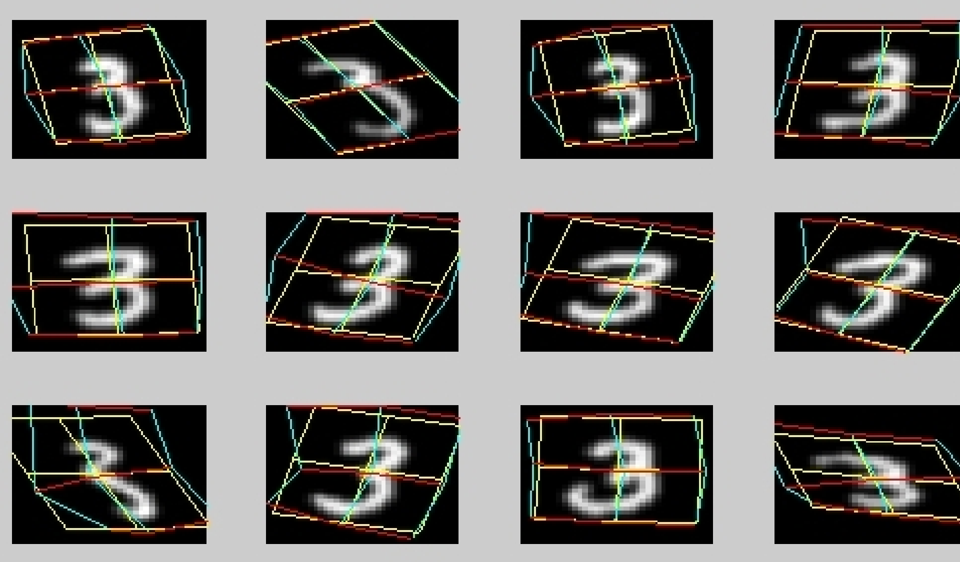


RASL-n

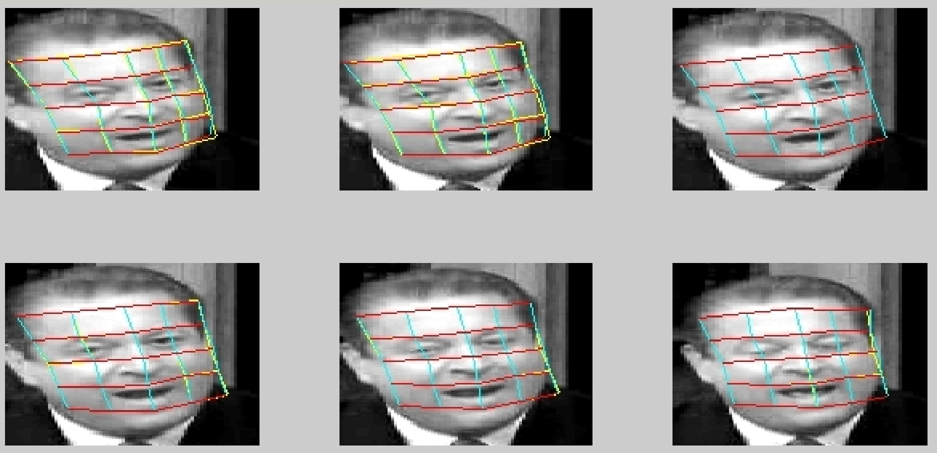


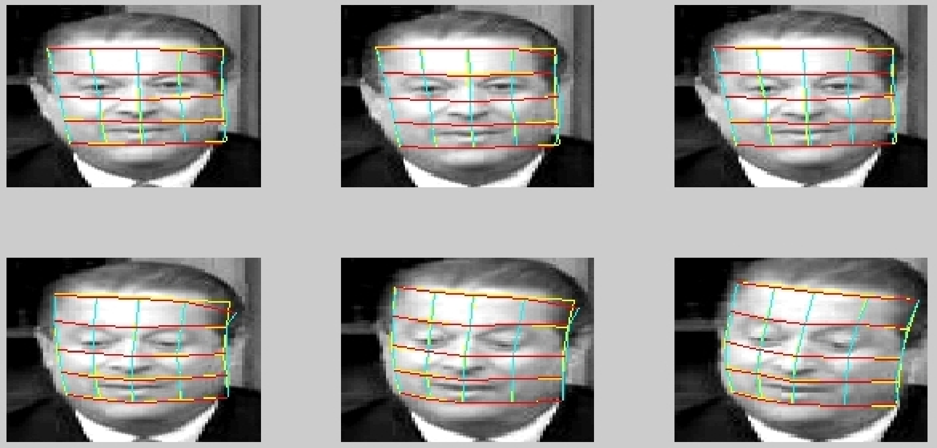
Low-rank A





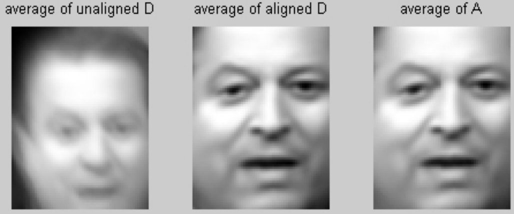
Gore



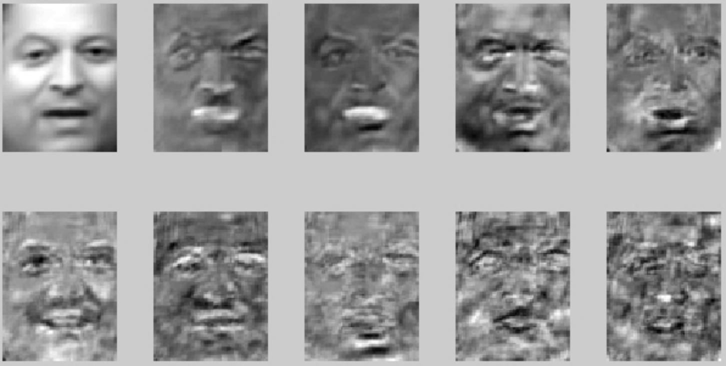


Aligned

Low-rank component



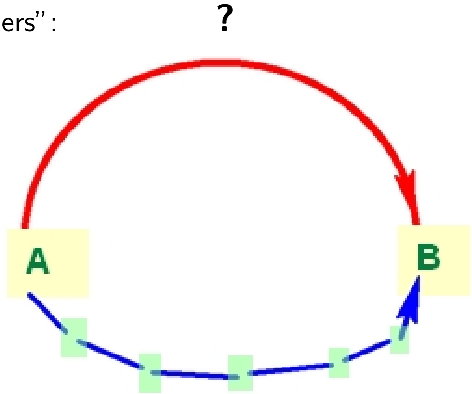
average image: before and after



Eigenfaces

Why RASL works

“Safety in numbers”:



Conclusions

* Has shown potential for improvement to RASL for aligning batches of images

* However, the method requires more tuning parameters to be chosen; is somewhat brittle + time consuming.

* Additional structure in video can be exploited via penalizing the difference in neighboring frames alignment parameters $\xi_t - \xi_{t+1}$ (in progress).

Future work

* more stable/ faster implementation of penalized RASL

* incorporate features of LDDMM

* online methods

* background/foreground

Thank you!

Reference

Yigang Peng, Arvind Ganesh, John Wright, Wenli Xu and Yi Ma

RASL: Robust Alignment via Sparse and Low-Rank Decomposition for Linearly Correlated Images (2012)

`http://www.columbia.edu/~jw2966/Peng12-PAMI.pdf`