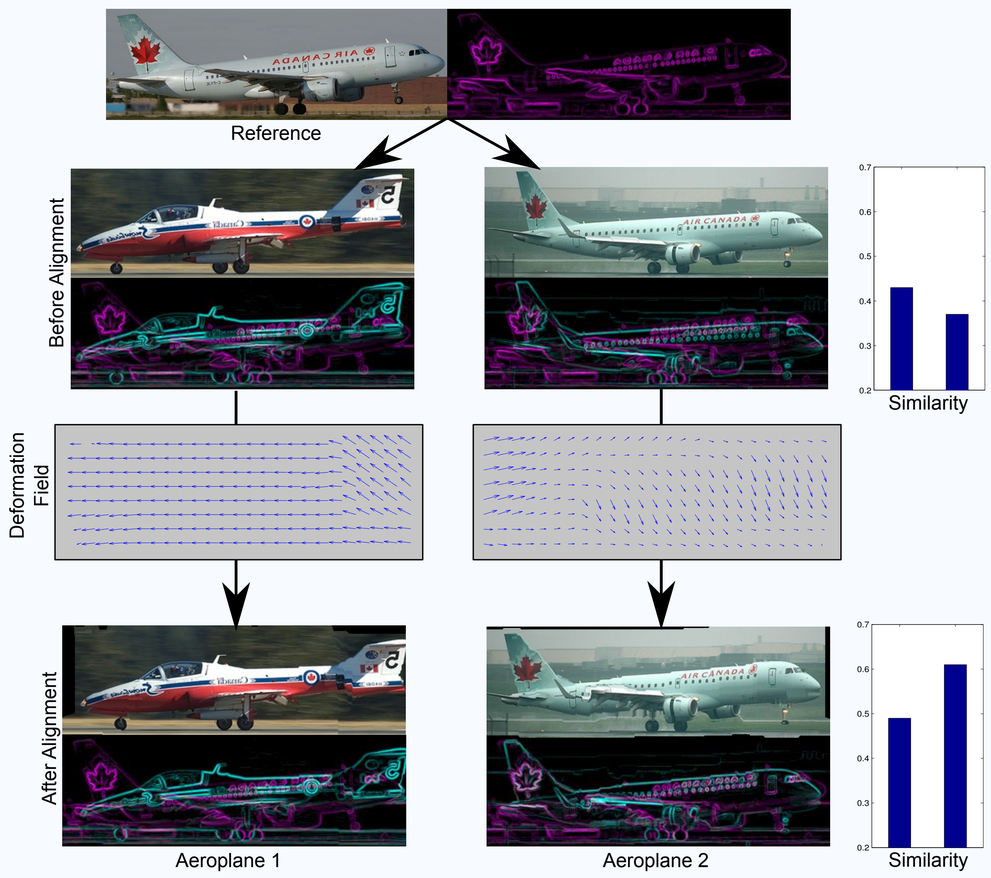
Distances based on Non-rigid Alignment for Comparison of Different Object Instances

Benjamin Drayer and Thomas Brox

Department of Computer Science and
BIOSS Centre for Biological Signalling Studies, University of Freiburg

Motivation

Distances based on alignment are more meaningful



Alignment

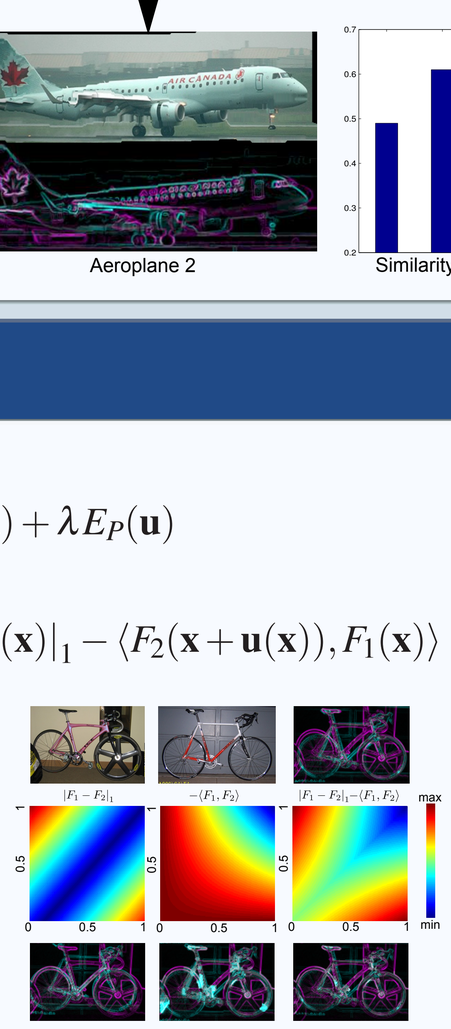
Minimize the energy:

$$E(\mathbf{u}) = E_D(\mathbf{u}) + \lambda E_P(\mathbf{u})$$

Data term:

$$E_D(\mathbf{u}) = \sum_{\mathbf{x}} |F_2(\mathbf{x}+\mathbf{u}(\mathbf{x})) - F_1(\mathbf{x})|_1 - \langle F_2(\mathbf{x}+\mathbf{u}(\mathbf{x})), F_1(\mathbf{x})\rangle$$

Combination of l_1-norm and dot product:



- l_1-norm: is robust but is likely to match weak features to the background
- dot product: Aligns all features but no direct penalty for unmatched features

Regularization term:

$$E_P(\mathbf{u}) = \sum_{\mathbf{x},\mathbf{y}\in\mathcal{N}(\mathbf{x})} |\mathbf{u}(\mathbf{x}) - \mathbf{u}(\mathbf{y})|_1,$$

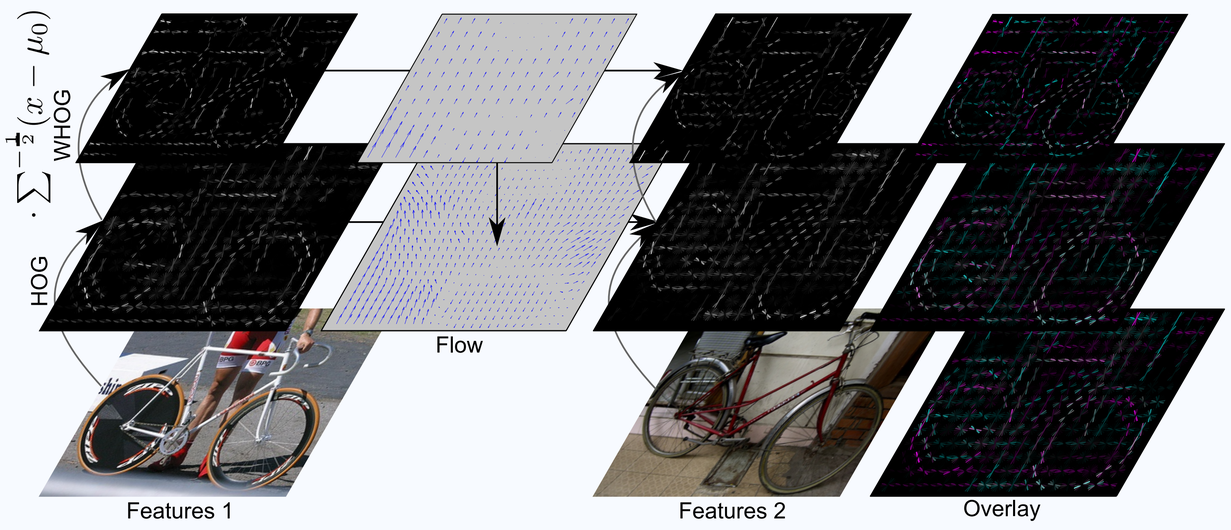
where $\mathcal{N}(\mathbf{x})$ denotes the neighborhood of $\mathbf{x}$.

Coarse to fine approach:

- Alignment on whitened HOG (WHOG) features [1]
  + Features more discriminative - Resolution is limited by construction
- Alignment on HOG features
  + Higher resolution $\Rightarrow$ finer details - More clutter
- Coarse to fine: Use the coarse deformation $\mathbf{u}_{\text{WHO}}$ as constraint for alignment on HOG level

$$E(\mathbf{u}) = \sum_{\mathbf{x}} \delta(\mathbf{x}) |\mathbf{u}_{\text{WHO}} - \mathbf{u}|_1 + E_D(\mathbf{u}) + \lambda E_P(\mathbf{u})$$

$$\delta(\mathbf{x}) = \begin{cases} 1, & \text{If } \mathbf{u}_{\text{WHO}} \text{ defined at } \mathbf{x} \\ 0, & \text{otherwise} \end{cases}$$

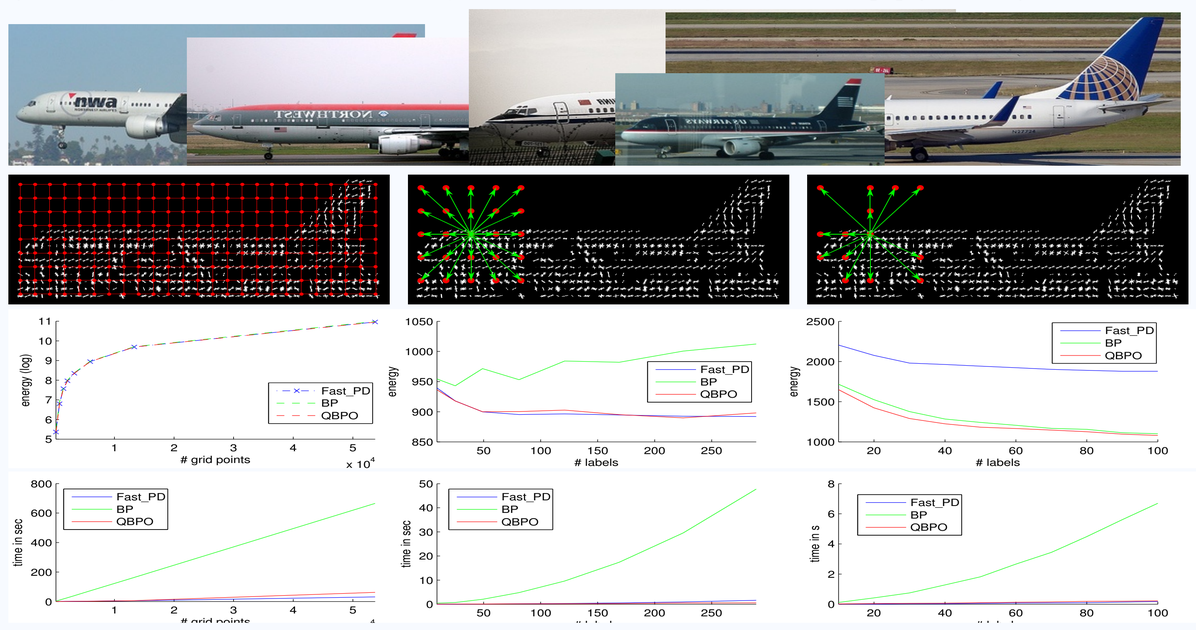


Optimization

Comparison between 3 state of the art techniques:

(1) Fast_PD [2] (2) Belief propagation (BP)

(3) $\alpha-$expansion with Quadratic pseudo boolean optimization [3] (QPBO)

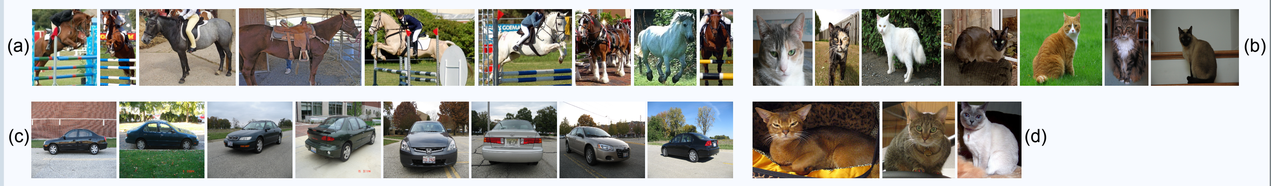


$\Rightarrow$ In our case α-expansion with QPBO (better scaling with a large amount of labels)

Results

Datasets:

Images show the different poses we defined for each dataset. (a) Horses from PASCAL 2007 (724) (b) Own cat dataset from Flickr (120), (c) 3D car dataset (80), (d) cats from PASCAL 2006 (200)



Evaluation:

For each pose we compute the precision and recall on the nearest neighbors (ground truth was manually labeled). For the feature based distances, we use

$$d(F_1, F_2) = \frac{\langle F_1, F_2\rangle}{\|F_1\|_2 \cdot \|F_2\|_2}.$$

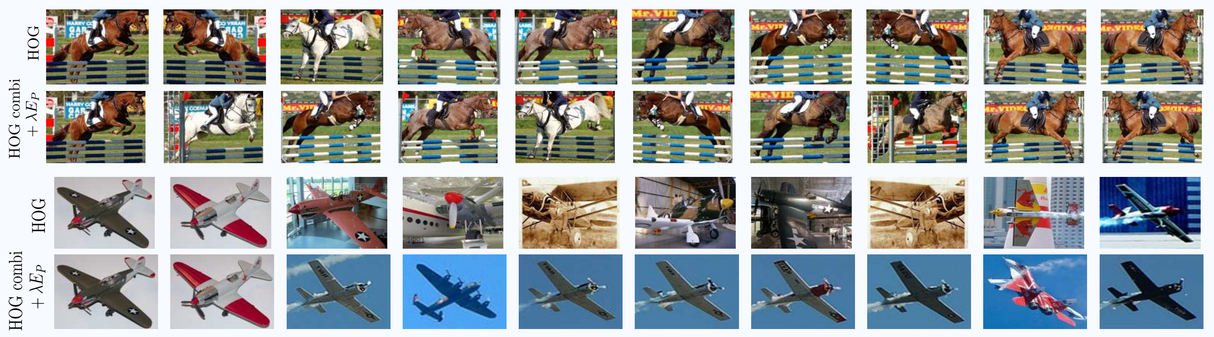
Table: Comparison of various distances with and without non-rigid alignment in terms of average precision (AP). The left block uses pure energies, the two blocks in the middle use HOG and WHO features, before and after the alignment. Methods with $+\lambda E_P$ make use of the deformation cost. The last two blocks use the coarse to fine method, which yields the best results.

	E_{HOG}	E_{WHO}	E_{combi}	HOG	HOG aligned	HOG $+\lambda E_P$	WHOG	WHO aligned	WHO $+\lambda E_P$	HOG combi	WHO combi	HOG combi $+\lambda E_P$	WHO combi $+\lambda E_P$
Cars	30.39	30.4	31.74	30.79	44.94	43.57	28.09	30.05	30.45	39.42	30.05	**46.01**	33.8
Cats own	13.6	13.78	13.81	31.18	31.97	32.23	33.04	30.66	30.99	32.93	30.66	**33.41**	32.29
Cats Pascal	6.55	6.61	6.56	33.16	31.81	31.05	31.32	30.97	31.24	27.82	27.42	32.58	**33.17**
Horse Pascal	4.32	4.31	4.22	29.49	36.47	37.38	33.34	35.43	36.42	36.87	35.43	**38.83**	33.8
Mean	13.72	13.78	14.08	31.16	36.3	36.1	31.45	31.78	32.28	34.26	30.89	**37.71**	33.27

Table: Performance of exemplar-SVM [4], rigid alignment and non-rigid aligned HOG-features. The non-rigid alignment consistently shows better AP.

	Cars	Cats own	Cats Pascal	Horse Pascal	Mean
ESVM [4]	24.07	16.83	10.68	19.08	17.67
rigid alignment	41.25	27.76	27.05	29.77	31.46
HOG aligned	44.94	31.97	31.81	36.47	**36.3**

For the reference images (left), we show the 9 nearest neighbors



[1] Bharath Hariharan, Jitendra Malik, and Deva Ramanan. Discriminative decorrelation for clustering and classification. In *ECCV*, 2012.

[2] Nikos Komodakis, Georgios Tziritas, and Nikos Paragios. Performance vs computational efficiency for optimizing single and dynamic mrfs: Setting the state of the art with primal-dual strategies. *Comput. Vis. Image Underst.*, 112(1):14–29, October 2008.

[3] Carsten Rother, Vladimir Kolmogorov, Victor Lempitsky, and Martin Szummer. Optimizing binary mrfs via extended roof duality. In *CVPR*, 2007.

[4] Tomasz Malisiewicz, Abhinav Gupta, and Alexei A. Efros. Ensemble of exemplar-svms for object detection and beyond. In *ICCV*, 2011.

We gratefully acknowledge funding by the Excellence Initiative of the German Federal and State Governments (EXC 294)