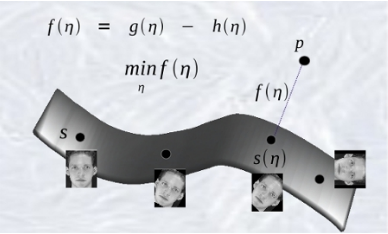
Geometry-aware analysis of high-dimensional visual information sets

Efi Kokiopoulou

PhD thesis at: Signal Processing Laboratory (LTS4), EPFL Lausanne
Supervisor: Prof. Pascal Frossard
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
Currently at: Seminar for Applied Mathematics, ETH Zürich

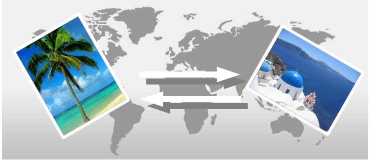
ACM Multimedia 2010



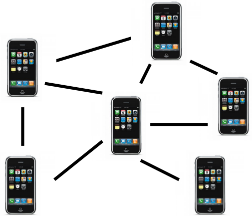
Motivation

Recent years:

- YouTube, Flickr, Picasa ...



- multimedia architectures: mobile devices, vision sensor networks ...

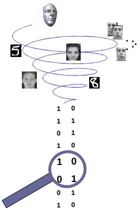


Explosion of multimedia data that need to be analyzed!

Challenges

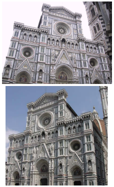
Such multimedia data:

need to be efficiently stored and analysed



can be geometrically transformed

captured in multiple observations



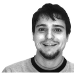
processed distributively

We study the classification of visual patterns in relation to the above challenges

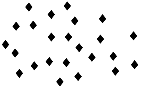
Illustrative application: face recognition

y_1 y_2 $\cdots$ y_{n-1} y_n

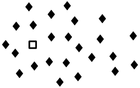
$y_t = ?$



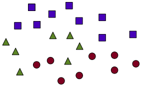
Training set



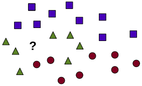
New data sample



Training set with class labels



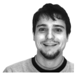
Needs to be classified



Illustrative application: face recognition

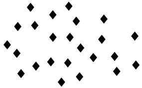
y_1 y_2 $\cdots$ y_{n-1} y_n

$y_t = ?$

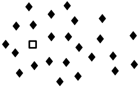


feature extraction

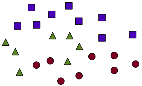
Training set



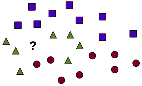
New data sample



Training set with class labels



Needs to be classified



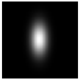
Geometric features

Dictionary: redundant set of basis functions (aka atoms)

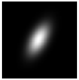
Parametric dictionary: produced by geometric transformations on a mother function

Sparse image representation: only a few basis functions are needed (K is small)

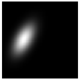
$\approx \alpha_1$



$+ \alpha_2$



$+ \ldots \quad \alpha_K$



Simultaneous sparse image representations: many images are approximated from the same set of atoms S

Approximation with set S

Flexible dimensionality reduction[1]

Goal: efficient data representation in

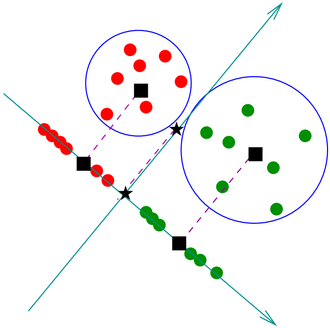
- storage
- classification

Semantic coding framework:

- greedy algorithm
- builds on simultaneous sparse representations
- parametric dictionary offers storage efficiency
- modified classification-aware criterion

$$J_{\text{approx}} + \lambda J_{\text{classif}}$$

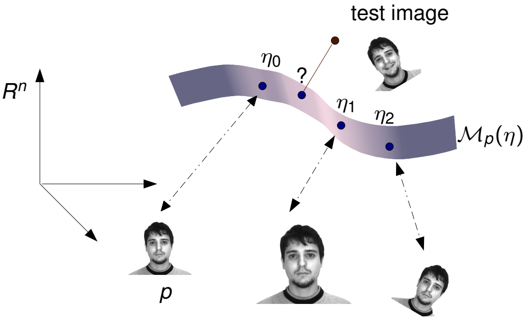
- λ: trade-off between approximation (storage) and classification performance



[1] E. Kokiopoulou and P. Frossard, "Semantic coding by supervised dimensionality reduction", IEEE Trans. on Multimedia, vol. 10, no 5, pp. 806-818, August 2008

Handling transformations

- Ideally, the feature representation should be adaptive to transformations
  - Meaningful comparison requires image alignment first
- Transformation manifold: set of all transformed versions of an image
- Each point on the manifold corresponds to a transformation η



Aligning a test image becomes equivalent to finding its closest point on the manifold

Problem formulation[2]

Image alignment: $\text{Minimize} \quad f(\eta) = \text{dist}(q, \mathcal{M}_p(\eta))$

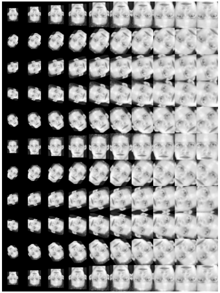
- hard non-convex optimization problem

Assumptions:

1. training image is sparsely approximated over a parametric dictionary
2. transformation consists of **rotation**, **translation** and **(isotropic) scaling** (4 parameters)

Thus:

- $f(\eta)$: closed form expression

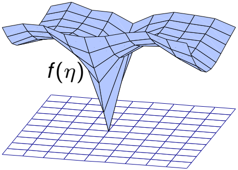


Theorem

The objective function $f(\eta)$ of the image alignment problem is ***DC*** *(difference of convex functions):*

$$f(\eta) = g(\eta) - h(\eta)$$

where g and h are convex functions.



[2] E. Kokiopoulou and P. Frossard, "Minimum distance between pattern transformation manifolds: Algorithm and Applications", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 7, pp. 1225-1238, July 2009.

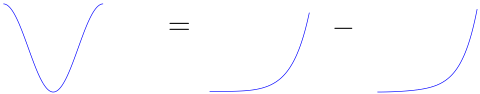
Globally optimal alignment via DC programming

Quick example: $\cos\theta,\ \theta \in [0, 2\pi]$

Globally optimal alignment via DC programming

Quick example: $\cos\theta,\ \theta \in [0, 2\pi]$

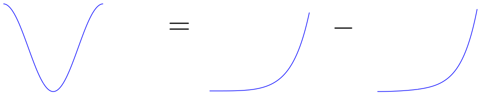
$$\cos\theta = \sum_{n=0}^{\infty} \frac{(-1)^n \theta^{2n}}{(2n)!} = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \ldots = \left[1 + \frac{\theta^4}{4!} + \ldots\right] - \left[\frac{\theta^2}{2!} + \frac{\theta^6}{6!} + \ldots\right]$$



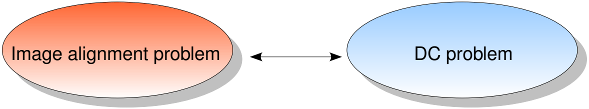
Globally optimal alignment via DC programming

Quick example: $\cos\theta,\ \theta \in [0, 2\pi]$

$$\cos\theta = \sum_{n=0}^{\infty} \frac{(-1)^n \theta^{2n}}{(2n)!} = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \ldots = \left[1 + \frac{\theta^4}{4!} + \ldots\right] - \left[\frac{\theta^2}{2!} + \frac{\theta^6}{6!} + \ldots\right]$$



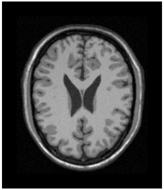
Overall:



- DC problems (despite non-convexity!) can be globally optimally solved
- Cutting plane method (R. Horst et al., '99)

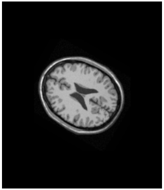
Our approach finds the global minimizer

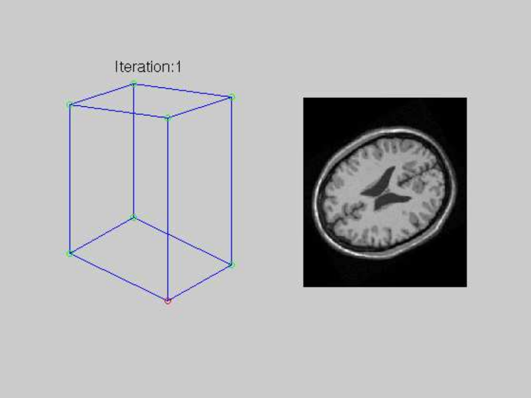
Demo



$\theta = 2\pi/3$

$\alpha = 0.6$





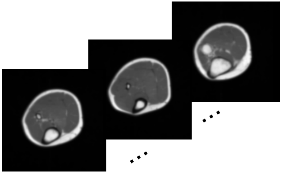
Applications

Face recognition robust to large transformations

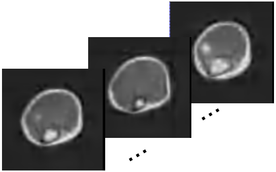
- effective classification thanks to the optimal alignment

Alignment of 3D volumetric images (joint work with N. Paragios and M. Zervos)

- 8 transformation parameters
- GPU implementation



original volume



sparse representation

Classification of multiple observations

Goal:

- classification of an object from a **set** of observations

Motivation:

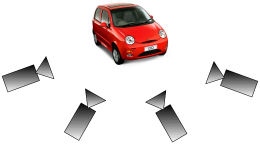
- diversity provides richer information

Typical applications:

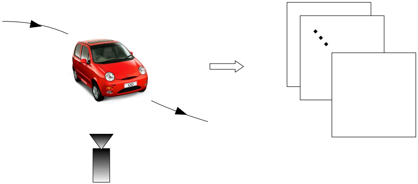
- multi-view object recognition
- video-based object recognition

Prior knowledge:

- manifold structure of the observations
- all observations belong to the same (unknown!) class

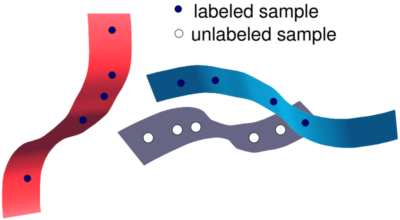


(a) multi-view object recognition



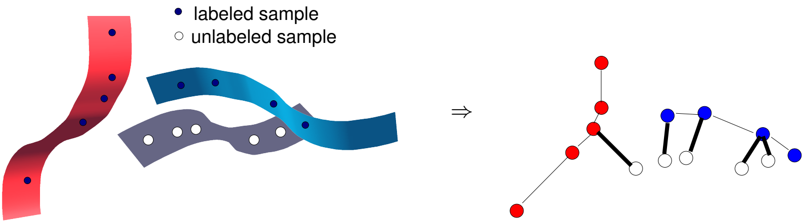
(b) video-based object recognition

Manifold classification



[3] E. Kokiopoulou and P. Frossard, “Graph-based classification of multiple observation sets”, Pattern Recognition, vol. 43(12), pp. 3988-3997, December 2010

Manifold classification

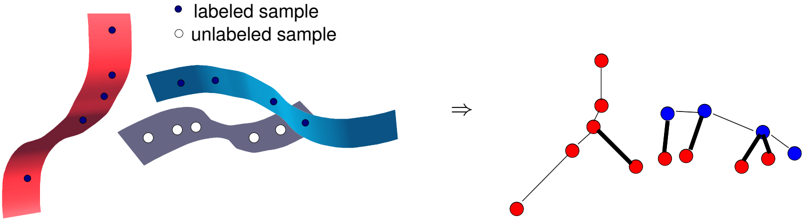


Graph-based classification algorithm[3]:

- based on the **smoothness assumption**: "*if two samples x_i and x_j are close-by, then it is likely that they share the same class label*"
- class hypothesis most consistent with the smoothness assumption is picked

[3] E. Kokiopoulou and P. Frossard, "Graph-based classification of multiple observation sets", Pattern Recognition, vol. 43(12), pp. 3988-3997, December 2010

Manifold classification

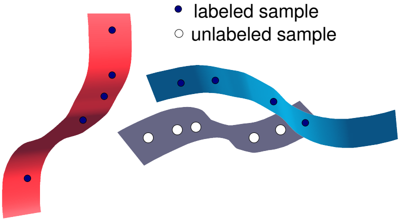


Graph-based classification algorithm[3]:

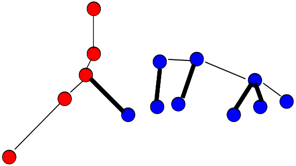
- based on the **smoothness assumption**: "*if two samples x_i and x_j are close-by, then it is likely that they share the same class label*"
- class hypothesis most consistent with the smoothness assumption is picked

[3] E. Kokiopoulou and P. Frossard, "Graph-based classification of multiple observation sets", Pattern Recognition, vol. 43(12), pp. 3988-3997, December 2010

Manifold classification



⇒



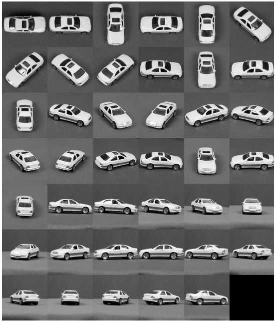
Graph-based classification algorithm[3]:

- based on the **smoothness assumption**: "*if two samples x_i and x_j are close-by, then it is likely that they share the same class label*"
- class hypothesis most consistent with the smoothness assumption is picked

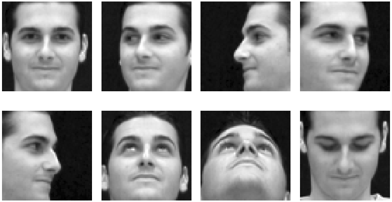
[3] E. Kokiopoulou and P. Frossard, "Graph-based classification of multiple observation sets", Pattern Recognition, vol. 43(12), pp. 3988-3997, December 2010

Illustrative applications

Multi-view object recognition
(ETH-80 data set)



Video face recognition
(VidTIMIT data set)

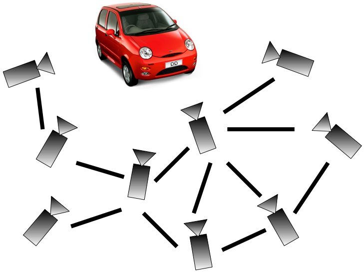


- Our method outperforms competing subspace or KLD methods

Distributed classification of multiple observations[4]

Distributed classification scenario:

- **ad-hoc** multimedia sensor networks without fusion center
- different observation at each sensor



Goal: to reach a common classification decision

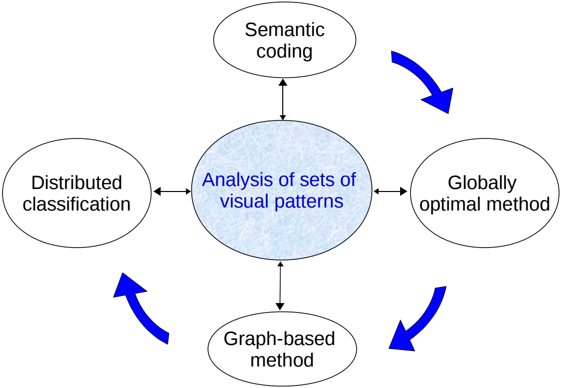
Distributed graph-based algorithm:

- local computation and communication
- all observations are progressively taken into account
- similar performance as a centralized solution

[4] E. Kokiopoulou and P. Frossard, "Distributed classification of multiple observation sets by consensus", IEEE Trans. on Signal Processing, in press.

E. Kokiopoulou and P. Frossard, "Polynomial filtering for fast convergence in distributed consensus", IEEE Transactions on Signal Processing, Vol. 57, Nr. 1, pp. 342-354, 2009.

Summary of thesis contributions



</TALK>

Thank you!

Questions?