Matrix and Tensor Factorizations in Vision

Christian Bauckhage, Georgios Evangelidis, Barbara Krausz, and Christian Thurau

universitätbonn b-it Fraunhofer IAIS

Bonn Vision Workshop, Oct. 8, 2009

Some Background Info

- roles:
  - media informatics @ B-IT/University of Bonn
  - multimedia pattern recognition @ Fraunhofer IAIS
- topics:
  - **multi-, mobile-, and social-media**
  - image/video retrieval and analysis
  - communities and web intelligence
  - game AI and agent behavior

NMF Activity Recognition
What?
How?
Application
Multilinear Discriminant Analysis
Basic Concepts
What? How?
Application
Archetypal Analysis
What?
Why?
How?
Application
Conclusion

Outline

Non-Negative Matrix Factorization for Activity Recognition
- What?
- How?
- Application

Multilinear Discriminant Analysis
- Basic Concepts
- What? How?
- Application

Archetypal Analysis of Large Image Sets
- What?
- Why?
- How?
- Application

Conclusion

Outline

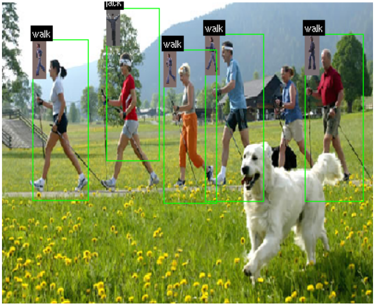
Non-Negative Matrix Factorization for Activity Recognition
- What?
- How?
- Application

Multilinear Discriminant Analysis
- Basic Concepts
- What? How?
- Application

Archetypal Analysis of Large Image Sets
- What?
- Why?
- How?
- Application

Conclusion

Goal: Activity Recognition in Images



- joint project with Vaclav Hlavac (CTU Prague) ...

Matrix Factorization

- given:

$$\mathbf{V}^{d \times n} \Leftrightarrow \text{data matrix}$$

- factorize s.t.

$$\mathbf{V} \approx \mathbf{W}\mathbf{H}$$

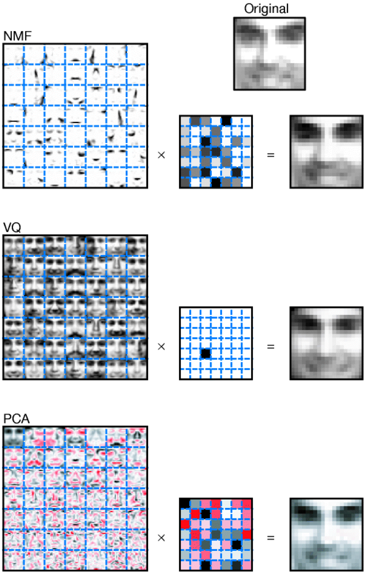
where

$$\mathbf{W}^{d \times k} \Leftrightarrow \text{basis vectors}$$
$$\mathbf{H}^{k \times n} \Leftrightarrow \text{coefficients}$$

- there may be constraints on **W** and **H**

Matrix Factorization; Common Constraints

- VQ:
  columns of **H** are unary vectors
- PCA:
  columns of **W** are orthonormal
- NMF:
  entries of **W** and **H** are non-negative



Non-negative Matrix Factorization

- determine **W** and **H** by minimizing least squares or KL-divergence[1][2]

$$\|\mathbf{V} - \mathbf{WH}\|^2 \quad \text{or} \quad D(\mathbf{V} \,\|\, \mathbf{WH})$$

- Update rule (shown for KL divergence):

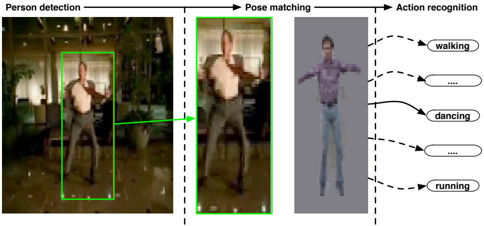
$$H_{i,j}^{t+1} \leftarrow {H^t}_{i,j} \frac{\sum_k W_{k,i} V_{i,j} / (WH)_{k,j}}{\sum_m W_{m,i}}$$

$$W_{j,k}^{t+1} \leftarrow {W^t}_{j,k} \frac{\sum_i H_{k,i} V_{j,i} / (WH)_{j,i}}{\sum_l H_{k,l}}$$

---

[1] [Lee, Nature'99]
[2] [Lee ,NIPS'01]

Application: Pose-based Action Recognition

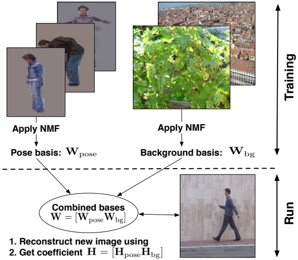


- idea: recognize activity based on a single pose
- applications: content based image retrieval, ...
- problem: pose estimation and background clutter

NMF Activity Recognition
What?
How?
Application
Multilinear Discriminant Analysis
Basic Concepts
What? How?
Application
Archetypal Analysis
What?
Why?
How?
Application
Conclusion

Application: Pose-based Action Recognition

- **idea:** decouple back- and foreground using NMF basis reconstruction
- apply NMF to *clean* human poses: basis vectors $\mathbf{W}_{pose}$
- apply NMF to background images: basis vectors $\mathbf{W}_{bg}$



Application: Pose-based Action Recognition

- estimate $\mathbf{W}_{pose}^{d \times k}$ and $\mathbf{W}_{bg}^{d \times k}$ during training, where

$$\mathbf{V}_{pose}^{d \times n} = \mathbf{W}_{pose}^{d \times k} \mathbf{H}_{pose}^{k \times n} \text{ and } \mathbf{V}_{bg}^{d \times m} = \mathbf{W}_{bg}^{d \times k} \mathbf{H}_{bg}^{k \times m}$$

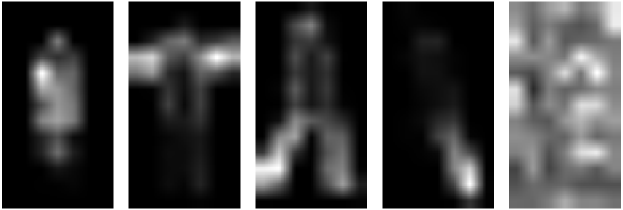
  where $\mathbf{H}_{pose}^{k \times n}$ can be interpreted as a pose descriptor
- For novel images $\mathbf{V}_{novel}^{d \times h}$ optimize for $\mathbf{H}_{novel}$ s.t.

$$\mathbf{V}_{novel}^{d \times h} = \mathbf{W}\mathbf{H}_{novel} = (\mathbf{W}_{pose}^{d \times k} \mathbf{W}_{bg}^{d \times k}) \begin{pmatrix} \mathbf{H}_{pose,novel}^{k \times h} \\ \mathbf{H}_{bg,novel}^{k \times h} \end{pmatrix}$$

- resulting $\mathbf{H}_{pose,novel}^{k \times h}$ describes pose and separates a foreground object from the background
- however, modeling arbitrary backgrounds is an ill posed problem

Application: Pose-based Action Recognition

NMF Activity Recognition
What?
How?
Application
Multilinear Discriminant Analysis
Basic Concepts
What? How?
Application
Archetypal Analysis
What?
Why?
How?
Application
Conclusion



- reconstruction of poses by parts
- coefficients $\mathbf{H}_{pose}$ encode the appearance of a pose (or better: a projection of $\mathbf{V}$ onto $\mathbf{W}_{pose}$)
- $\mathbf{W}_{pose}^{d \times k}$ and $\mathbf{W}_{bg}^{d \times k}$ enable generative detection

Generative Model for Human Detection

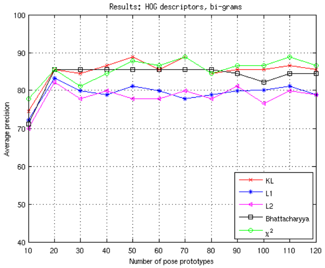
- **idea:** $\mathbf{W}_{pose}^{d \times k}$ and $\mathbf{W}_{bg}^{d \times k}$ generate $\mathbf{V}$ independently
- classify based on a likelihood ratio of independent models [Bissacco et al.,NIPS'06]
- $\mathbf{I} = \mathbf{V} \approx \mathbf{W}\mathbf{H}$, for combined bases $\mathbf{W} = [\mathbf{W}_{\mathbf{pose}} \mathbf{W}_{\mathbf{bg}}]$

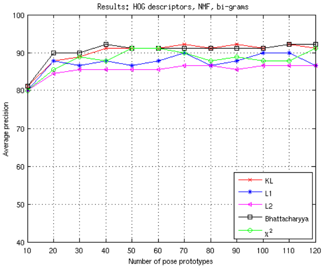
$$L = \frac{P(\mathbf{I}|bg)}{P(\mathbf{I}|pose)} \sim \frac{1 - |\mathbf{V} - \mathbf{V}_{\mathbf{pose}}|/|\mathbf{V}|}{1 - |\mathbf{V} - \mathbf{V}_{\mathbf{bg}}|/|\mathbf{V}|}. \tag{1}$$

- activity can be understood as a sequence of poses
- express activities as distributions over a set of pose primitives

Results on Weizmann Data Set

- 10 action performed by 9 subjects





Methods	(%) sequences	(%) still images
No background subtraction and applicable to single frames		
Thurau et al.[Dagstuhl'09]	93	70
Niebles et al. [CVPR'07]	72.8	55.0
Weinland et al. [CVPR'08]	93.6	-
Ferrari at al. [CVPR'08]	88.0	-
Thurau et al. [CVPR'08]	94.40	70.4

Outline

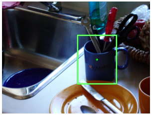
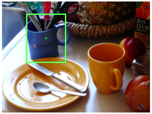
Non-Negative Matrix Factorization for Activity Recognition
What?
How?
Application

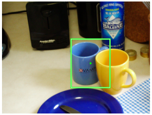
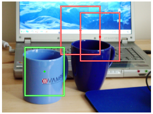
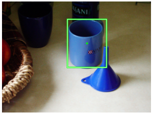
Multilinear Discriminant Analysis
Basic Concepts
What? How?
Application

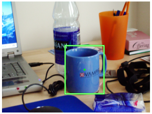
Archetypal Analysis of Large Image Sets
What?
Why?
How?
Application

Conclusion

Goal: Template Learning for Object Detection




Basic Concepts

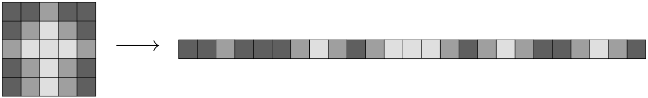
Image Data in Classification

observation:

- digital intensity images $\Leftrightarrow$ $m \times n$ arrays
- digital color images $\Leftrightarrow$ $m \times n \times 3$ arrays

common practice:

- representation as vectors in $\mathbb{R}^{mn}$ or $\mathbb{R}^{3mn}$
- example



(a) $\boldsymbol{X} = [X_{ij}]$ (b) $\mathbf{x}^T = [x_k]$, where $k = i \cdot n + j$

however …

NMF Activity Recognition
What?
How?
Application
Multilinear Discriminant Analysis
Basic Concepts
What? How?
Application
Archetypal Analysis
What?
Why?
How?
Application
Conclusion

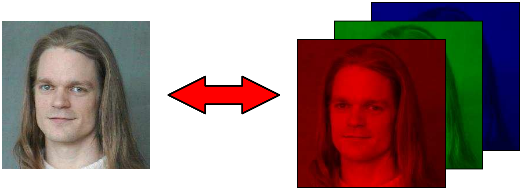
Basic Concepts
3rd Order Tensors

- color image $\Leftrightarrow$ 3rd order tensor

$$\boldsymbol{\mathcal{A}} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$$

- elements

$$\mathcal{A}_{ijk} \in \mathbb{R}$$



Basic Concepts

Some Tensor Algebra and Structures

- inner product

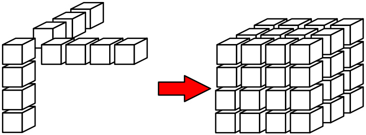
$$\mathcal{A} \cdot \mathcal{B} = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \sum_{k=1}^{m_n} \mathcal{A}_{ijk} \mathcal{B}_{ijk} \stackrel{!}{=} \mathcal{A}_{ijk} \mathcal{B}_{ijk}$$

- rank-1 tensor

$$\mathcal{A} = \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w}$$

i.e.

$$\mathcal{A}_{ijk} = u_i v_j w_k$$



Basic Concepts
Tensor Decompositions

- if there is

$$\mathcal{A}^* = \sum_r \mathbf{u}^r \otimes \mathbf{v}^r \otimes \mathbf{w}^r$$

such that

$$\mathcal{A}^* = \operatorname*{argmin}_{\mathcal{A}'} \left\| \mathcal{A} - \mathcal{A}' \right\|_F$$

$\mathcal{A}$ has a PARAFAC model

NMF Activity Recognition
What?
How?
Application
Multilinear Discriminant Analysis
Basic Concepts
What? How?
Application
Archetypal Analysis
What?
Why?
How?
Application
Conclusion

Multilinear Filter Design

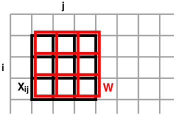
Runtime

filtering intensity images $\Leftrightarrow$ 2D correlation/convolution:

- no constraints

$$Y_{ij} = (\boldsymbol{I} * \boldsymbol{W})_{ij} = \sum_{m,n} X_{ij} W_{i-m,j-n}$$

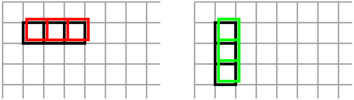
$\Longrightarrow O(mn)$ per pixel



- PARAFAC constraints

$$Y_{ij} = (\boldsymbol{I} * \boldsymbol{W})_{ij} = \sum_{r=1}^{\rho} \left(\boldsymbol{I} * \mathbf{u}_r\right) * \mathbf{v}_r$$

$\Longrightarrow O(\rho(m+n))$ per pixel



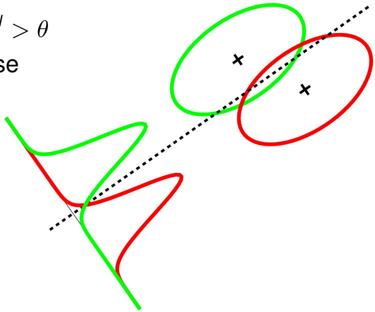
Multilinear Filter Design

Linear Discriminant Analysis

2 class problem:

- given: $\{\mathbf{x}^l, y^l\}_{l=1,\ldots,N}$
- find: projection **w** and threshold θ

$$\omega(\mathbf{x}^l) = \begin{cases} \omega_p, & \text{if} \quad \mathbf{w}^T\mathbf{x}^l > \theta \\ \omega_n, & \text{otherwise} \end{cases}$$



Fisher, 1936: 2 solutions

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmax}} \frac{\mathbf{w}^T\mathbf{S}_b\mathbf{w}}{\mathbf{w}^T\mathbf{S}_w\mathbf{w}}$$

$$\Leftrightarrow \mathbf{S}_b\mathbf{w} = \lambda\mathbf{S}_w\mathbf{w}$$

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmin}} \left(\mathbf{y} - \boldsymbol{X}\mathbf{w}\right)^T\left(\mathbf{y} - \boldsymbol{X}\mathbf{w}\right)$$

$$= \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\mathbf{y}$$

Multilinear Filter Design

Least Squares for LDA

NMF Activity Recognition
What?
How?
Application
Multilinear Discriminant Analysis
Basic Concepts
What? How?
Application
Archetypal Analysis
What?
Why?
How?
Application
Conclusion

- OLS solution

$$\mathbf{w} = \left(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathsf{T}}\mathbf{y} \tag{2}$$

  sensitive to noise/corrupted samples

- better: RLS solution

$$\mathbf{w} = \left(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X} + \lambda \boldsymbol{I}\right)^{-1}\boldsymbol{X}^{\mathsf{T}}\mathbf{y} \tag{3}$$

- better yet: KLS solution

$$\mathbf{w} = \boldsymbol{X}^{T}\left(\boldsymbol{K} + \lambda \boldsymbol{I}\right)^{-1}\mathbf{y} \tag{4}$$

  where, e.g., $K_{ij} = \exp\left(-\frac{\|\mathbf{x}^i - \mathbf{x}^j\|^2}{2\sigma^2}\right)$

Multilinear Filter Design

Tensor Discriminant Analysis

- consider tensorial least squares problem
- i.e. minimize

$$E(\mathcal{W}) = \sum_{l} \left(\mathcal{W} \cdot \mathcal{X}^{l} - y^{l} \right)^2 \tag{5}$$

- assume PARAFAC model for $\mathcal{W}$
- i.e. constrain $\mathcal{W}$ to

$$\mathcal{W} = \sum_{r=1}^{\rho} \mathbf{u}^{r} \otimes \mathbf{v}^{r} \otimes \mathbf{w}^{r}$$

↯ this precludes closed form solution to (5)

Multilinear Filter Design

Tensor Discriminant Analysis

NMF Activity Recognition
What?
How?
Application
Multilinear Discriminant Analysis
Basic Concepts
What? How?
Application
Archetypal Analysis
What?
Why?
How?
Application
Conclusion

- consider tensorial least squares problem
- i.e. minimize

$$E(\mathcal{W}) = \sum_{l} \left(\mathcal{W} \cdot \mathcal{X}^l - y^l \right)^2 \tag{5}$$

- assume PARAFAC model for $\mathcal{W}$
- i.e. constrain $\mathcal{W}$ to

$$\mathcal{W} = \sum_{r=1}^{\rho} \mathbf{u}^r \otimes \mathbf{v}^r \otimes \mathbf{w}^r$$

↯ this precludes closed form solution to (5)

Multilinear Filter Design

Alternating Least Squares Algorithm

Input: a training set $\{\mathcal{X}^l, y^l\}_{l=1,\ldots,N}$ of tensors $\mathcal{X}^l \in \mathbb{R}^{m_1 \times \cdots \times m_n}$ with class labels y^l
Output: a rank-ρ approximation of an nth-order projection tensor $\mathcal{W} = \mathbf{u}_1^r \otimes \mathbf{u}_2^r \otimes \ldots \otimes \mathbf{u}_n^r$

for $r = 1, \ldots, \rho$
  $t = 0$
  for $j = 1, \ldots, n-1$
    randomly initialize $\mathbf{u}_j^r(t)$
    orthogonalize $\mathbf{u}_j^r(t)$ w.r.t. $\{\mathbf{u}_j^1, \ldots, \mathbf{u}_j^{r-1}\}$
  repeat
    $t \leftarrow t+1$
    for $j = n, \ldots, 1$
      for $l = 1, \ldots, N$
        contract $x_{i_j}^l = \mathcal{X}^l_{i_1\ldots i_{j-1} i_j i_{j+1}\ldots i_n} u_{i_1}^r(t) \ \ldots \ u_{i_{j-1}}^r(t) \ u_{i_{j+1}}^r(t) \ \ldots \ u_{i_n}^r(t)$
      $\mathbf{u}_j^r(t) = \operatorname{argmin}_{\mathbf{u_j^r}} \|\boldsymbol{X}^T \mathbf{u}_j^r - \mathbf{y}\|^2$, where $\boldsymbol{X} = [\mathbf{x}_{i_j}^1, \ldots, \mathbf{x}_{i_j}^N]^T$ and $\mathbf{y} = [y^1, \ldots, y^N]^T$
      orthogonalize $\mathbf{u}_j^r(t)$ w.r.t. $\{\mathbf{u}_j^1, \ldots, \mathbf{u}_j^{r-1}\}$
  until $\|\mathbf{u}_1^r(t) - \mathbf{u}_1^r(t-1)\| \leq \epsilon \ \vee \ t > t_{\max}$
endfor

Multilinear Filter Design

Alternating Least Squares Algorithm

Input: a training set $\{\mathcal{X}^l, y^l\}_{l=1,\ldots,N}$ of tensors $\mathcal{X}^l \in \mathbb{R}^{m_1 \times \cdots \times m_n}$ with class labels y^l
Output: a rank-ρ approximation of an nth-order projection tensor $\mathcal{W} = \mathbf{u}_1^r \otimes \mathbf{u}_2^r \otimes \ldots \otimes \mathbf{u}_n^r$

for $r = 1, \ldots, \rho$
  $t = 0$
  for $j = 1, \ldots, n-1$
    randomly initialize $\mathbf{u}_j^r(t)$
    orthogonalize $\mathbf{u}_j^r(t)$ w.r.t. $\{\mathbf{u}_j^1, \ldots, \mathbf{u}_j^{r-1}\}$
  repeat
    $t \leftarrow t + 1$
    for $j = n, \ldots, 1$
      for $l = 1, \ldots, N$
        contract $x_{i_j}^l = \mathcal{X}^l_{i_1 \ldots i_{j-1} i_j i_{j+1} \ldots i_n} u^r_{i_1}(t) \ \ldots \ u^r_{i_{j-1}}(t) \ u^r_{i_{j+1}}(t) \ \ldots \ u^r_{i_n}(t)$
      $\mathbf{u}_j^r(t) = \operatorname{argmin}_{\mathbf{u}_j^r} \|\boldsymbol{X}^T \mathbf{u}_j^r - \mathbf{y}\|^2$, where $\boldsymbol{X} = [\mathbf{x}_{i_j}^1, \ldots, \mathbf{x}_{i_j}^N]^T$ and $\mathbf{y} = [y^1, \ldots, y^N]^T$
      orthogonalize $\mathbf{u}_j^r(t)$ w.r.t. $\{\mathbf{u}_j^1, \ldots, \mathbf{u}_j^{r-1}\}$
  until $\|\mathbf{u}_1^r(t) - \mathbf{u}_1^r(t-1)\| \leq \epsilon \ \vee \ t > t_{\max}$
endfor

Multilinear Filter Design

Training Time

- rapid training

$$\left.\begin{aligned} \mathbf{u}^T\mathbf{u} &: m_1 \times m_1 \\ \mathbf{v}^T\mathbf{v} &: m_2 \times m_2 \\ \mathbf{w}^T\mathbf{w} &: m_3 \times m_3 \end{aligned}\right\} \ll \boldsymbol{X}^T\boldsymbol{X} : m_1 m_2 m_3 \times m_1 m_2 m_3$$

- robustness against *small sample sizes*
- example: car detection [Bauckhage, Käster, ICPR'06]

NMF Activity Recognition
What?
How?
Application
Multilinear Discriminant Analysis
Basic Concepts
What? How?
Application
Archetypal Analysis
What?
Why?
How?
Application
Conclusion

Multilinear Filter Design

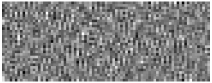
Performance

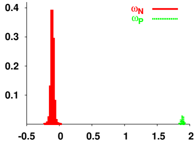
NMF Activity Recognition
What?
How?
Application
Multilinear Discriminant Analysis
Basic Concepts
What? How?
Application
Archetypal Analysis
What?
Why?
How?
Application
Conclusion

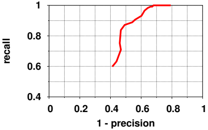
example: car detection (cntd.)

- $\mathbf{w} = \boldsymbol{X}^{\dagger}\mathbf{y}$

  training: 35 s
  test: 11 s

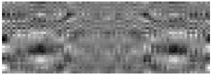


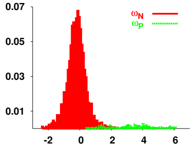


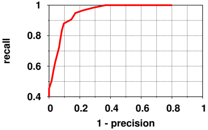


- $\mathbf{W} = \sum_r \mathbf{u}_r \otimes \mathbf{v}_r$

  training: 5 s
  test: 4 s







Application

Results (Template Learning, AR Face Data)

OLS

(a) trained with 3 samples (b) trained with 6 samples

RLS

(c) trained with 3 samples (d) trained with 6 samples

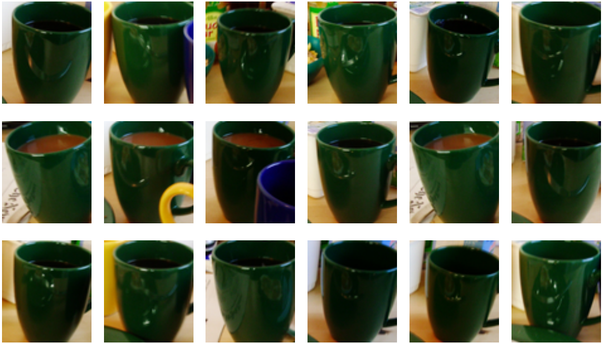
KLS

(e) trained with 3 samples (f) trained with 6 samples

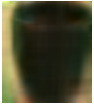
NMF Activity Recognition
What?
How?
Application
Multilinear Discriminant Analysis
Basic Concepts
What? How?
Application
Archetypal Analysis
What?
Why?
How?
Application
Conclusion

Application

Results (Template Learning, Cup Data)



⋮

OLS  **RLS**  **KLS** 

NMF Activity Recognition
What?
How?
Application
Multilinear Discriminant Analysis
Basic Concepts
What? How?
Application
Archetypal Analysis
What?
Why?
How?
Application
Conclusion

Experiments

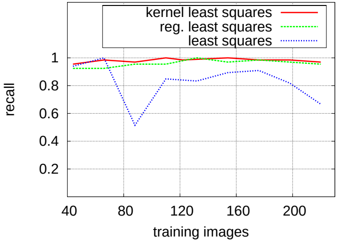
Setting (Object Detection)

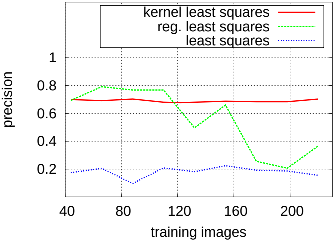
- dataset: breakfast scene collection
  - 22 training images
  - item 66 test images
- target: green cup
  - 22 unaligned patches of a cup
  - up to 198 counter examples
  - size: $91 \times 71 \times 3$
  - $\rho = 6$
- 2 designs
  1. one stage detection
  2. two stage detection

Experiments

Results (Object Detection)

- recall and precision (one stage detection)





NMF Activity Recognition
What?
How?
Application
Multilinear Discriminant Analysis
Basic Concepts
What? How?
Application
Archetypal Analysis
What?
Why?
How?
Application
Conclusion

Experiments

Results (Object Detection)

- on the importance of high precision . . .

method	recall	precision
one stage OLS	1.00	0.20
two stage KLS	0.98	1.00

Outline

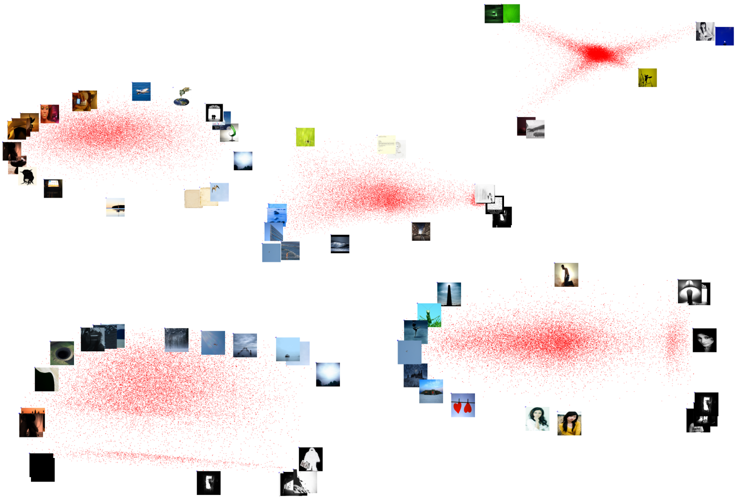
Non-Negative Matrix Factorization for Activity Recognition
- What?
- How?
- Application

Multilinear Discriminant Analysis
- Basic Concepts
- What? How?
- Application

Archetypal Analysis of Large Image Sets
- What?
- Why?
- How?
- Application

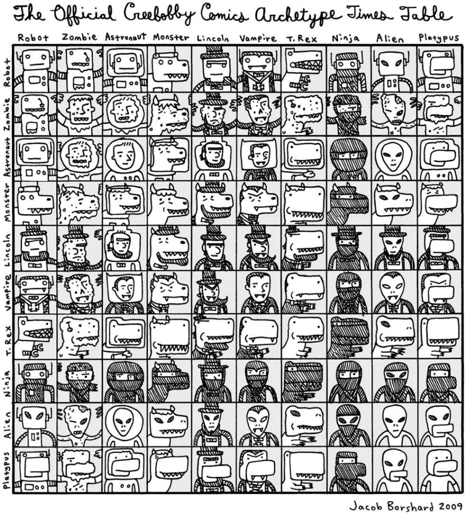
Conclusion

Goal: Identify Structures in Image Collections



NMF Activity Recognition
What?
How?
Application
Multilinear Discriminant Analysis
Basic Concepts
What? How?
Application
Archetypal Analysis
What?
Why?
How?
Application
Conclusion

Archetypes

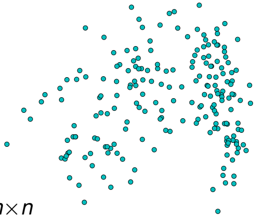


- **Plato**:
  ideals; pure forms that embody fundamental characteristics of a thing rather than its specific peculiarities
- **C.G. Jung**:
  innate, universal forms (the *hero*, the *great mother*, the *wise old man*, . . .) that channel experiences and emotions, resulting in recognizable and typical behaviors with certain probable outcomes
- **A. Cutler** and **L. Breiman** (in Technometrics 36(4), 1994):
  archetypal analysis $\Longleftrightarrow$ new way of data analysis for multivariate data

NMF Activity Recognition
What?
How?
Application
Multilinear Discriminant Analysis
Basic Concepts
What? How?
Application
Archetypal Analysis
What?
Why?
How?
Application
Conclusion

What?

what is archetypal analysis?



- AA assumes a data matrix

$$\boldsymbol{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n] \in \mathbb{R}^{m \times n}$$

and considers a constrained optimization problem

$$\text{RSS}(p) = \min_{\boldsymbol{A},\boldsymbol{B}} \left\| \boldsymbol{X} - \boldsymbol{X}\boldsymbol{B}\boldsymbol{A} \right\|^2$$

where

$$\boldsymbol{A} \in \mathbb{R}^{p \times n}, \ \boldsymbol{A} \succeq \mathbf{0}, \ \sum_{k=1}^{p} a_{kl} = 1, \quad l = 1, \ldots, n$$

$$\boldsymbol{B} \in \mathbb{R}^{n \times p}, \ \boldsymbol{B} \succeq \mathbf{0}, \ \sum_{j=1}^{n} b_{jl} = 1, \quad l = 1, \ldots, p$$

What?

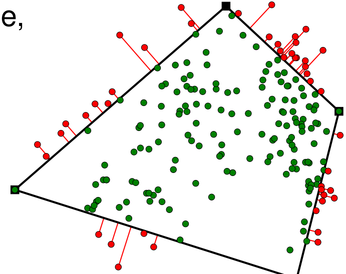
what is archetypal analysis?

- substituting $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{B} \in \mathbb{R}^{m \times p}$ (usually $p \ll n$) yields

$$\mathbf{z}_k = \sum_{j=1}^{n} \mathbf{x}_j b_{jk} \qquad \text{and} \qquad \left\| \mathbf{x}_i - \sum_{k=1}^{p} \mathbf{z}_k a_{ki} \right\|^2$$

$\Leftrightarrow$ the *archetypes* $\mathbf{z}_k$ are sparse, convex mixtures of the data $\mathbf{x}_i$

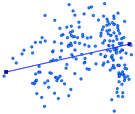
$\Leftrightarrow$ the data $\mathbf{x}_i$ are sparse, convex mixtures of archetypes $\mathbf{z}_k$



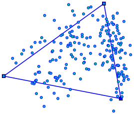
What?

what is archetypal analysis?

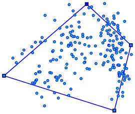
- archetypes provably reside on the data convex hull
- increasing p approximates the hull



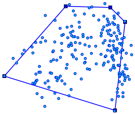
$p = 2$



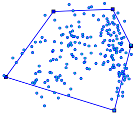
$p = 3$



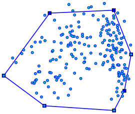
$p = 4$



$p = 5$

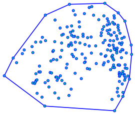


$p = 6$



$p = 7$

$\dots$



convex hull

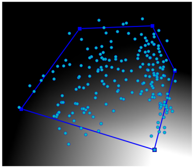
Why?

why is it interesting?

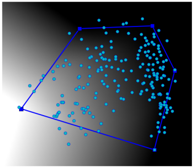
- recall: $\mathbf{x}_i = \boldsymbol{Z}\mathbf{a}_i$ with stochastic coefficient vectors $\mathbf{a}_i$

$\Rightarrow$ the coefficients a_{ki} can be thought of as $P(\mathbf{x}_i | \mathbf{z}_k)$

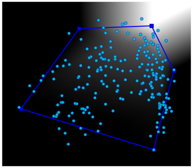
$\Rightarrow$ (soft)clustering, classification, ranking, . . .



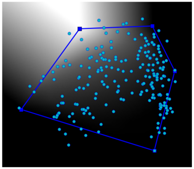
$P(\mathbf{x}|\mathbf{z}_1)$



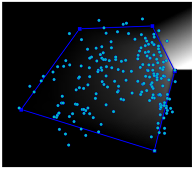
$P(\mathbf{x}|\mathbf{z}_2)$



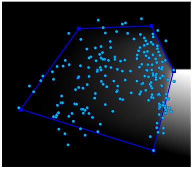
$P(\mathbf{x}|\mathbf{z}_3)$



$P(\mathbf{x}|\mathbf{z}_4)$



$P(\mathbf{x}|\mathbf{z}_5)$



$P(\mathbf{x}|\mathbf{z}_6)$

How?

how is it computed?

- **algorithm according to Cutler and Breiman**:

1. determine coefficients a_{ki} by solving n constrained problems $\min \left\| \boldsymbol{Z} \mathbf{a}_i - \mathbf{x}_i \right\|^2$ s.t. $a_{ki} \geq 0$ and $\sum_k a_{ki} = 1$
2. given the updated a_{ki}, compute intermediate archetypes

$$\tilde{\boldsymbol{Z}} = \boldsymbol{X} \boldsymbol{A}^T \left(\boldsymbol{A} \boldsymbol{A}^T \right)^{-1}$$

3. determine coefficients b_{jk} by solving p constrained problems $\min \left\| \boldsymbol{X} \mathbf{b}_k - \tilde{\mathbf{z}}_k \right\|^2$ s.t. $b_{jk} \geq 0$ and $\sum_j b_{jk} = 1$
4. update the archetypes by setting $\boldsymbol{Z} = \boldsymbol{X} \boldsymbol{B}$
5. compute the new RSS; unless it falls below a threshold or only marginally improves the old RSS, continue at 1.

NMF Activity Recognition
What?
How?
Application
Multilinear Discriminant Analysis
Basic Concepts
What? How?
Application
Archetypal Analysis
What?
Why?
How?
Application
Conclusion

Fraunhofer IAIS

How?

how is it computed?

- **algorithm according to Cutler and Breiman**:

1. determine coefficients a_{ki} by solving n constrained problems $\min \left\| \boldsymbol{Z}\mathbf{a}_i - \mathbf{x}_i \right\|^2$ s.t. $a_{ki} \geq 0$ and $\sum_k a_{ki} = 1$
2. given the updated a_{ki}, compute intermediate archetypes
$$\tilde{\boldsymbol{Z}} = \boldsymbol{X}\boldsymbol{A}^T \left(\boldsymbol{A}\boldsymbol{A}^T\right)^{-1}$$
3. determine coefficients b_{jk} by solving p constrained problems $\min \left\| \boldsymbol{X}\mathbf{b}_k - \tilde{\mathbf{z}}_k \right\|^2$ s.t. $b_{jk} \geq 0$ and $\sum_j b_{jk} = 1$
4. update the archetypes by setting $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{B}$
5. compute the new RSS; unless it falls below a threshold or only marginally improves the old RSS, continue at 1.

NMF Activity Recognition
What?
How?
Application
Multilinear Discriminant Analysis
Basic Concepts
What? How?
Application
Archetypal Analysis
What?
Why?
How?
Application
Conclusion

How?

how is it computed?

- **algorithm according to Cutler and Breiman**:

1. determine coefficients a_{ki} by solving n constrained problems $\min \left\| \boldsymbol{Z}\mathbf{a}_i - \mathbf{x}_i \right\|^2$ s.t. $a_{ki} \geq 0$ and $\sum_k a_{ki} = 1$
2. given the updated a_{ki}, compute intermediate archetypes

$$\tilde{\boldsymbol{Z}} = \boldsymbol{X}\boldsymbol{A}^T \left(\boldsymbol{A}\boldsymbol{A}^T\right)^{-1}$$

3. determine coefficients b_{jk} by solving p constrained problems $\min \left\| \boldsymbol{X}\mathbf{b}_k - \tilde{\mathbf{z}}_k \right\|^2$ s.t. $b_{jk} \geq 0$ and $\sum_j b_{jk} = 1$
4. update the archetypes by setting $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{B}$
5. compute the new RSS; unless it falls below a threshold or only marginally improves the old RSS, continue at 1.

Fraunhofer IAIS

How?

how is it computed?

- **algorithm according to Cutler and Breiman**:

1. determine coefficients a_{ki} by solving n constrained problems $\min \left\| \boldsymbol{Z}\mathbf{a}_i - \mathbf{x}_i \right\|^2$ s.t. $a_{ki} \geq 0$ and $\sum_k a_{ki} = 1$
2. given the updated a_{ki}, compute intermediate archetypes

$$\tilde{\boldsymbol{Z}} = \boldsymbol{X}\boldsymbol{A}^T \left(\boldsymbol{A}\boldsymbol{A}^T\right)^{-1}$$

3. determine coefficients b_{jk} by solving p constrained problems $\min \left\| \boldsymbol{X}\mathbf{b}_k - \tilde{\mathbf{z}}_k \right\|^2$ s.t. $b_{jk} \geq 0$ and $\sum_j b_{jk} = 1$
4. update the archetypes by setting $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{B}$
5. compute the new RSS; unless it falls below a threshold or only marginally improves the old RSS, continue at 1.

How?

how is it computed?

- **algorithm according to Cutler and Breiman**:

1. determine coefficients a_{ki} by solving n constrained problems min $\left\|\boldsymbol{Z}\mathbf{a}_i - \mathbf{x}_i\right\|^2$ s.t. $a_{ki} \geq 0$ and $\sum_k a_{ki} = 1$
2. given the updated a_{ki}, compute intermediate archetypes

$$\tilde{\boldsymbol{Z}} = \boldsymbol{X}\boldsymbol{A}^T\left(\boldsymbol{A}\boldsymbol{A}^T\right)^{-1}$$

3. determine coefficients b_{jk} by solving p constrained problems min $\left\|\boldsymbol{X}\mathbf{b}_k - \tilde{\mathbf{z}}_k\right\|^2$ s.t. $b_{jk} \geq 0$ and $\sum_j b_{jk} = 1$
4. update the archetypes by setting $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{B}$
5. compute the new RSS; unless it falls below a threshold or only marginally improves the old RSS, continue at 1.

NMF Activity Recognition
What?
How?
Application
Multilinear Discriminant Analysis
Basic Concepts
What? How?
Application
Archetypal Analysis
What?
Why?
How?
Application
Conclusion

How?

how is it computed?

- **algorithm according to Cutler and Breiman**:

1. determine coefficients a_{ki} by solving n constrained problems min $\left\|\boldsymbol{Z}\mathbf{a}_i - \mathbf{x}_i\right\|^2$ s.t. $a_{ki} \geq 0$ and $\sum_k a_{ki} = 1$
2. given the updated a_{ki}, compute intermediate archetypes

$$\tilde{\boldsymbol{Z}} = \boldsymbol{X}\boldsymbol{A}^T\left(\boldsymbol{A}\boldsymbol{A}^T\right)^{-1}$$

3. determine coefficients b_{jk} by solving p constrained problems min $\left\|\boldsymbol{X}\mathbf{b}_k - \tilde{\mathbf{z}}_k\right\|^2$ s.t. $b_{jk} \geq 0$ and $\sum_j b_{jk} = 1$
4. update the archetypes by setting $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{B}$
5. compute the new RSS; unless it falls below a threshold or only marginally improves the old RSS, continue at 1.

How?

how is it computed?

- analysis:

- in step 1: $i = 1, \ldots, n$ problems involving matrices of size p^2
- in step 3: $k = 1, \ldots, p$ problems

$$\begin{aligned} \min \; & \frac{1}{2}\mathbf{b}_j^T \boldsymbol{R}\mathbf{b}_k - \mathbf{r}^T\mathbf{b}_k, \qquad \boldsymbol{R} = \boldsymbol{X}^T\boldsymbol{X} \in \mathbb{R}^{n \times n}, \;\; \mathbf{r} = \boldsymbol{X}^T\tilde{\mathbf{z}}_k \in \mathbb{R}^n \\ \text{s.t. } & \boldsymbol{I}\,\mathbf{b}_k \geq \mathbf{0} \\ & \mathbf{1}^T\mathbf{b}_k = 1 \end{aligned}$$

- recall: p = number of archetypes: n = number of data points

$\Rightarrow$ step 3 involves matrices of size n^2 and costs dearly

How?

how is it computed?

- analysis:
- in step 1: $i = 1, \ldots, n$ problems involving matrices of size p^2
- in step 3: $k = 1, \ldots, p$ problems

$$\begin{aligned} \min \ & \frac{1}{2}\mathbf{b}_j^T \boldsymbol{R}\mathbf{b}_k - \mathbf{r}^T\mathbf{b}_k, \qquad \boldsymbol{R} = \boldsymbol{X}^T\boldsymbol{X} \in \mathbb{R}^{n\times n}, \ \ \mathbf{r} = \boldsymbol{X}^T\tilde{\mathbf{z}}_k \in \mathbb{R}^n \\ \text{s.t. } & \boldsymbol{I}\,\mathbf{b}_k \geq \mathbf{0} \\ & \mathbf{1}^T\mathbf{b}_k = 1 \end{aligned}$$

- recall: $p =$ number of archetypes: $n =$ number of data points

$\Rightarrow$ step 3 involves matrices of size n^2 and costs dearly

How?

how can it be made practical?

- **improvement (I): working sets**

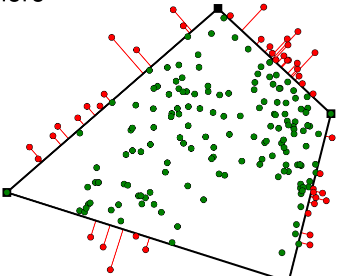
- in each iteration, consider $X = X^+ \cup X^-$ where

$$X^- = \{\mathbf{x}_i \in X | \mathbf{x}_i = \boldsymbol{Z}\mathbf{a}_i\}$$
$$X^+ = \{\mathbf{x}_i \in X | \mathbf{x}_i \neq \boldsymbol{Z}\mathbf{a}_i\}$$

- hence $\boldsymbol{X} = \left[\boldsymbol{X}^+ \, \boldsymbol{X}^-\right]$ where

$$\boldsymbol{X}^+ \in \mathbb{R}^{m \times n'}$$
$$\boldsymbol{X}^- \in \mathbb{R}^{m \times (n-n')}$$

and $n' < n$



Fraunhofer IAIS

How?

how can it be made practical?

- this yields:

$$\left\|\boldsymbol{X} - \boldsymbol{Z}\boldsymbol{A}\right\|^2 = \left\|\left[\boldsymbol{X}^+ \, \boldsymbol{X}^-\right] - \boldsymbol{Z}\left[\boldsymbol{A}^+ \, \boldsymbol{A}^-\right]\right\|^2$$

$$= \underbrace{\left\|\boldsymbol{X}^+ - \boldsymbol{Z}\boldsymbol{A}^+\right\|^2}_{\neq 0} + \underbrace{\left\|\boldsymbol{X}^- - \boldsymbol{Z}\boldsymbol{A}^-\right\|^2}_{=0}$$

and with $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{B}$, where $\boldsymbol{B}^- = \boldsymbol{0}$, it further reduces to

$$\left\|\boldsymbol{X}^+ - \boldsymbol{Z}\boldsymbol{A}^+\right\|^2 = \left\|\boldsymbol{X}^+ - \left[\boldsymbol{X}^+ \, \boldsymbol{X}^-\right]\begin{bmatrix}\boldsymbol{B}^+ \\ \boldsymbol{B}^-\end{bmatrix}\boldsymbol{A}^+\right\|^2$$

$$= \left\|\boldsymbol{X}^+ - \boldsymbol{X}^+\boldsymbol{B}^+\boldsymbol{A}^+\right\|^2$$

- effort in step 3 reduces to $O(n'^2) < O(n^2)$

NMF Activity Recognition
What?
How?
Application
Multilinear Discriminant Analysis
Basic Concepts
What? How?
Application
Archetypal Analysis
What?
Why?
How?
Application
Conclusion

How?

how can it be made practical?

- this yields:

$$\left\| \boldsymbol{X} - \boldsymbol{Z}\boldsymbol{A} \right\|^2 = \left\| \left[\boldsymbol{X}^+ \, \boldsymbol{X}^- \right] - \boldsymbol{Z} \left[\boldsymbol{A}^+ \, \boldsymbol{A}^- \right] \right\|^2$$

$$= \underbrace{\left\| \boldsymbol{X}^+ - \boldsymbol{Z}\boldsymbol{A}^+ \right\|^2}_{\neq 0} + \underbrace{\left\| \boldsymbol{X}^- - \boldsymbol{Z}\boldsymbol{A}^- \right\|^2}_{= 0}$$

and with $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{B}$, where $\boldsymbol{B}^- = \boldsymbol{0}$, it further reduces to

$$\left\| \boldsymbol{X}^+ - \boldsymbol{Z}\boldsymbol{A}^+ \right\|^2 = \left\| \boldsymbol{X}^+ - \left[\boldsymbol{X}^+ \, \boldsymbol{X}^- \right] \begin{bmatrix} \boldsymbol{B}^+ \\ \boldsymbol{B}^- \end{bmatrix} \boldsymbol{A}^+ \right\|^2$$

$$= \left\| \boldsymbol{X}^+ - \boldsymbol{X}^+ \boldsymbol{B}^+ \boldsymbol{A}^+ \right\|^2$$

- effort in step 3 reduces to $O(n'^2) < O(n^2)$

How?

how can it be made practical?

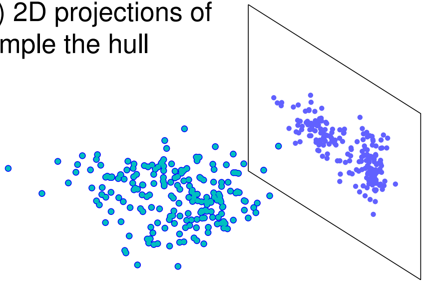
- **improvement (II): sampling the convex hull**

- archetypes are mixtures of points on the data convex hull

$\Rightarrow$ restrict algorithm to $X^H \subseteq X$

- in $\mathbb{R}^m$, convex hull computation is "expensive" ($\Theta(n^{(m/2)})$)

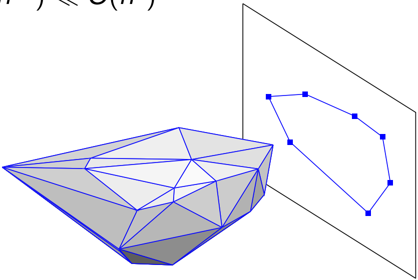
$\Rightarrow$ consider (many) 2D projections of the data and sample the hull



How?

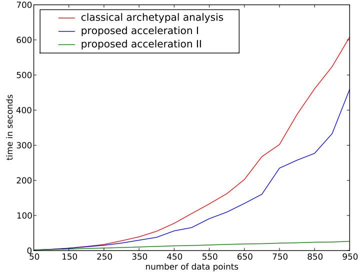
how can it be made practical?

- every image of a polytope P under an affine map $\pi : \mathbf{x} \rightarrow \boldsymbol{M}\mathbf{x} + \mathbf{t}$ is a polytope
- in particular, every vertex of an affine image of P corresponds to a vertex of P
- sampling the hull is “cheap”
- effort is then $O(n''^2) \ll O(n^2)$
- n'' is $\Omega(\sqrt{\log n})$



How?

how can it be made practical?



Input: data matrix $\boldsymbol{X} \in \mathbb{R}^{m \times n}$
Output: matrix of archetypes $\boldsymbol{Z} \in \mathbb{R}^{m \times p}$
and coefficient matrices $\boldsymbol{A}$ and $\boldsymbol{B}$

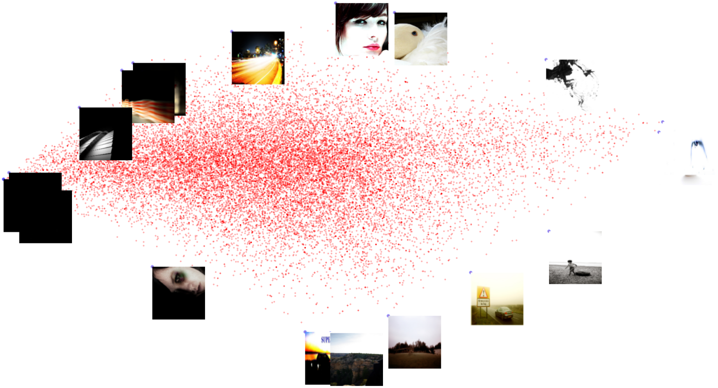
---

preselect archetypal candidates $\boldsymbol{X}^H$
initialize matrices $\boldsymbol{Z}$, $\boldsymbol{A}$, and $\boldsymbol{B}$
compute $\text{RSS}_{t=0}$
repeat
 optimize $\boldsymbol{A} = \min_{\boldsymbol{A}} \|\boldsymbol{X}^H - \boldsymbol{Z}\boldsymbol{A}\|^2$
  s.t. $a_{ji} \geq 0$ and $\sum_j a_{ji} = 1$
 determine working set X^+
 determine matrices $\boldsymbol{X}^+$, $\boldsymbol{A}^+$, and $\tilde{\boldsymbol{Z}}^+$
 set $\boldsymbol{B}^- = \boldsymbol{0}$
 optimize $\boldsymbol{B}^+ = \min_{\boldsymbol{B}^+} \|\tilde{\boldsymbol{Z}}^+ - \boldsymbol{X}^+\boldsymbol{B}^+\|^2$
  s.t. $b_{ij} \geq 0$ and $\sum_i b_{ij} = 1$
 update the archetypes $\boldsymbol{Z} = \boldsymbol{X}^+\boldsymbol{B}^+$
until $\text{RSS}_{t+1} < \theta$ or $|\text{RSS}_{t+1} - \text{RSS}_t| < \epsilon$

NMF Activity Recognition
What?
How?
Application
Multilinear Discriminant Analysis
Basic Concepts
What? How?
Application
Archetypal Analysis
What?
Why?
How?
Application
Conclusion

Application

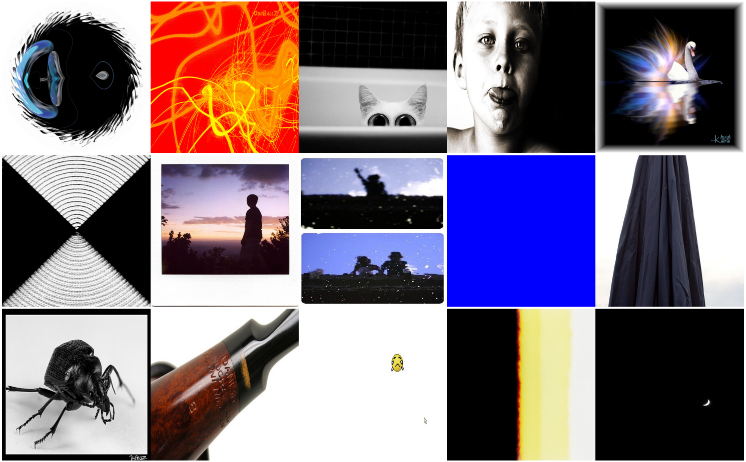
A convex hull projections of 50.000 flickr images



Fraunhofer IAIS

Application

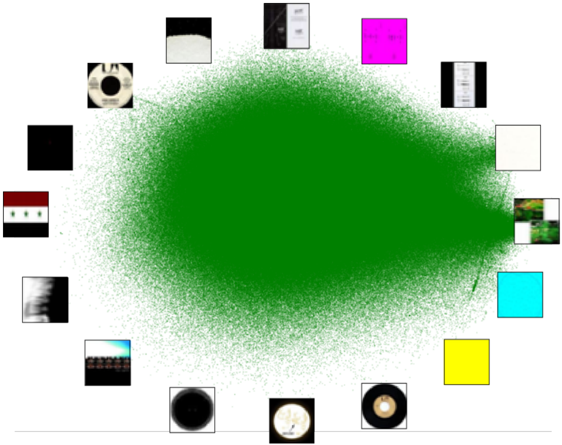
The archetypes of 50.000 flickr Images



Application

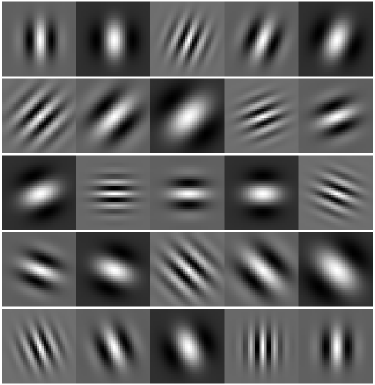
1.600.000 tiny images provided by Torralba, Fergus and Weiss

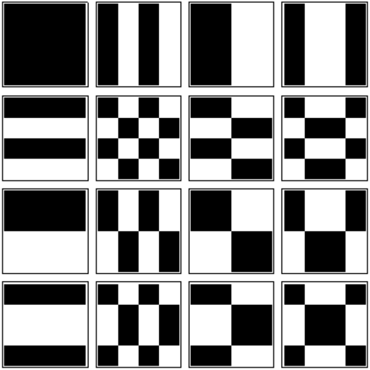
- 3072 dim. RGB color features
- 16 archetypal images



Application

One cannot help but notice . . .





Outline

Non-Negative Matrix Factorization for Activity Recognition
What?
How?
Application

Multilinear Discriminant Analysis
Basic Concepts
What? How?
Application

Archetypal Analysis of Large Image Sets
What?
Why?
How?
Application

Conclusion

Conclusion I

- Data representations by matrix factorization: $\mathbf{V} \approx \mathbf{WH}$
- Reconstruction using independently generated basis vectors: $\mathbf{W}_1, \mathbf{W}_2$
- Applied for pose estimation in cluttered images
- So far: sparseness in **W** and **H** for conic combinations
- Next: further constraints to **W** and **H** for efficient pattern indexing

Conclusion II

- tensor-based approach to filter design
- incorporating the kernel-trick increases
  - robustness under presence of outliers
  - robustness under high degree of data variability

Conclusion III

- AA is an interesting, “novel” approach to data analysis and classification
- exploiting its geometry drastically accelerates AA so that it becomes *practical*
- caveat:
    - cases where $m \gg n$