Non-negative Matrix Factorization for Multimodal Image Retrieval

Fabio A. González PhD

Bioingenium Research Group
Computer Systems and Industrial Engineering Department
Universidad Nacional de Colombia
Knowledge Discovery and Webmining Lab
CECS, The U of L

Graduate Seminar, Spring 2010
CECS Department , The University of Louisville

Outline

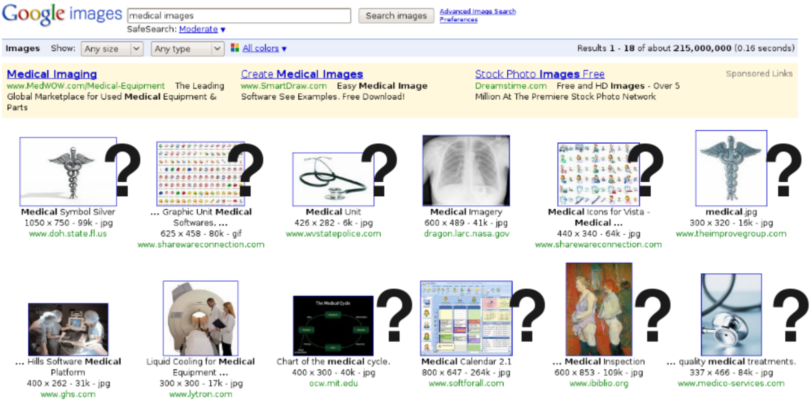
1. The Problem
   - Content-based image retrieval
   - Semantic image retrieval
   - Multimodal image retrieval
2. Non-Negative Matrix Factorization
   - The Netflix Prize
   - Non-negative matrix factorization
   - NMF vs SVD
3. NMF for Multimodal Retrieval
   - Semantic space
   - Multimodal clustering
   - Image annotation
   - Multimodal retrieval

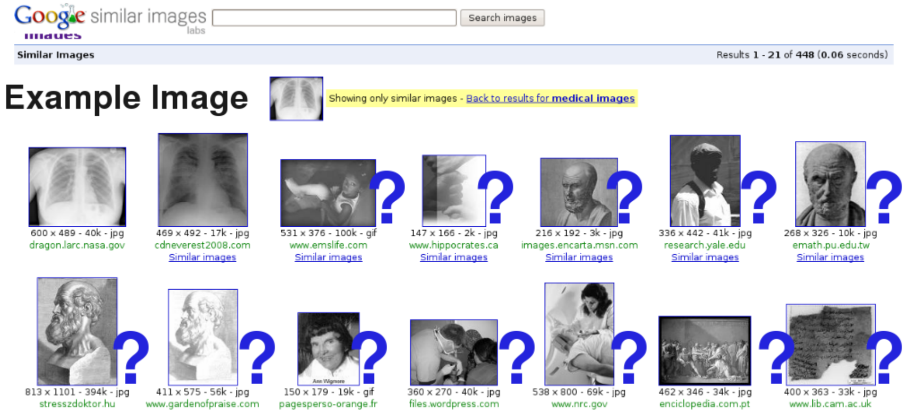
Outline

Outline

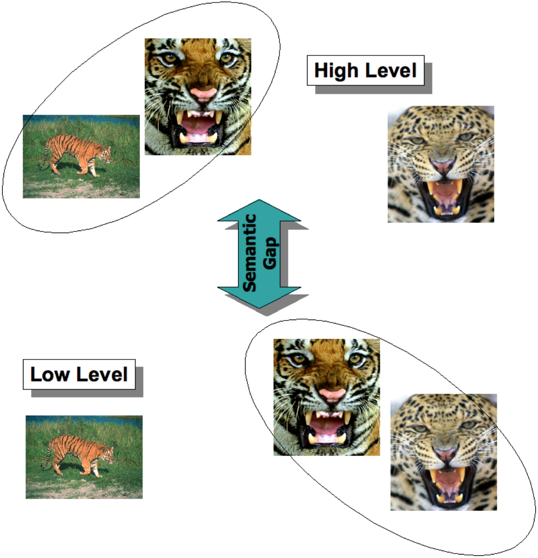
1. The Problem
   - Content-based image retrieval
   - Semantic image retrieval
   - Multimodal image retrieval
2. Non-Negative Matrix Factorization
   - The Netflix Prize
   - Non-negative matrix factorization
   - NMF vs SVD
3. NMF for Multimodal Retrieval
   - Semantic space
   - Multimodal clustering
   - Image annotation
   - Multimodal retrieval

Content-Based Image Retrieval



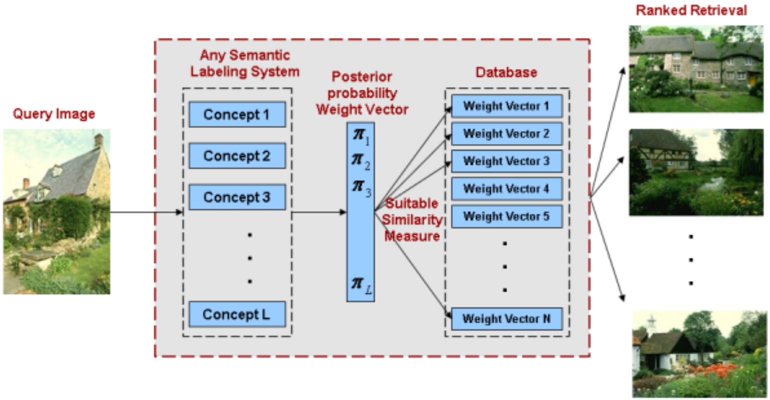

Google similar images labs
Search images
Similar Images
Results 1 - 21 of 448 (0.06 seconds)
Example Image
Showing only similar images - Back to results for medical images
600 x 489 - 40k - jpg
dragon.larc.nasa.gov
469 x 492 - 17k - jpg
cdneverest2008.com
Similar images
531 x 376 - 100k - gif
www.emslife.com
Similar images
147 x 166 - 2k - jpg
www.hippocrates.ca
Similar images
216 x 192 - 3k - jpg
images.encarta.msn.com
Similar images
336 x 442 - 41k - jpg
research.yale.edu
Similar images
268 x 326 - 10k - jpg
emath.pu.edu.tw
Similar images
813 x 1101 - 394k - jpg
stresszdoktor.hu
411 x 575 - 56k - jpg
www.gardenofpraise.com
150 x 179 - 19k - gif
pagesperso-orange.fr
360 x 270 - 40k - jpg
files.wordpress.com
538 x 800 - 69k - jpg
www.nrc.gov
462 x 346 - 34k - jpg
enciclopedia.com.pt
400 x 363 - 33k - jpg
www.lib.cam.ac.uk

Low-level vs. High-level



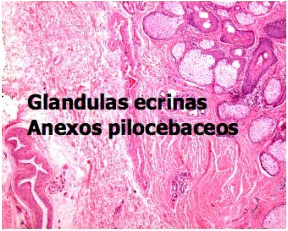
Outline

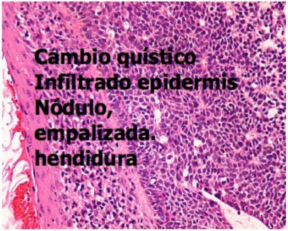
Semantic Annotation using ML

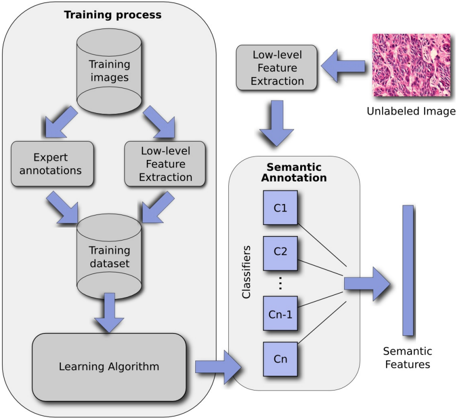


Source: Nuno Vasconcelos, UCSD, http://www.svcl.ucsd.edu/projects/qbse/

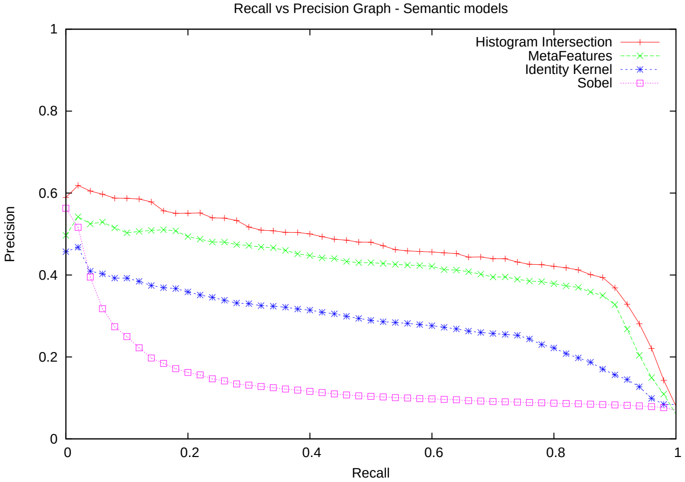
An Example (1)


Glandulas ecrinas
Anexos pilocebaceos


Cambio quístico
Infiltrado epidermis
Nódulo,
empalizada.
hendidura


Training process
Training images
Expert annotations
Low-level Feature Extraction
Training dataset
Learning Algorithm
Low-level Feature Extraction
Unlabeled Image
Semantic Annotation
Classifiers
C1
C2
Cn-1
Cn
Semantic Features

An Example (2)



Recall vs Precision Graph - Semantic models
Histogram Intersection
MetaFeatures
Identity Kernel
Sobel
Precision
Recall
0
0.2
0.4
0.6
0.8
1

Disadvantages

- Requires a training set with expert annotations, so it is a costly process

Disadvantages

- Requires a training set with expert annotations, so it is a costly process
- Does not scale for large semantic vocabularies

Disadvantages

- Requires a training set with expert annotations, so it is a costly process
- Does not scale for large semantic vocabularies
- The mapping from visual features to annotations may lose the visual richness

Outline

1. The Problem
   - Content-based image retrieval
   - Semantic image retrieval
   - Multimodal image retrieval
2. Non-Negative Matrix Factorization
   - The Netflix Prize
   - Non-negative matrix factorization
   - NMF vs SVD
3. NMF for Multimodal Retrieval
   - Semantic space
   - Multimodal clustering
   - Image annotation
   - Multimodal retrieval

Multimodality

Text and images come naturally together in many documents

- Academic papers, books
- Newspapers, web pages
- Medical cases

Multimodal Retrieval

- Unstructured text associated to images may be used as semantic annotations
  - Images and texts are complimentary information units
  - Take advantage of interactions between both data modalities

Multimodal Retrieval

- Unstructured text associated to images may be used as semantic annotations
  - Images and texts are complimentary information units
  - Take advantage of interactions between both data modalities
- Problems:
  - Text associated to images is not structured
  - Unclear relationships between keywords and visual patterns
  - Possible presence of noise in both data modalities

Multimodal Retrieval

- Unstructured text associated to images may be used as semantic annotations
  - Images and texts are complimentary information units
  - Take advantage of interactions between both data modalities
- Problems:
  - Text associated to images is not structured
  - Unclear relationships between keywords and visual patterns
  - Possible presence of noise in both data modalities
- Retrieval scenarios:
  - Cross-modal:
    - find images based on a text query
    - find text based on an image query (image annotation)
  - Visual retrieval based on a visual query

Outline

Outline

The Netflix Competition

- Problem: prediction of user ratings for films (collaborative filtering)

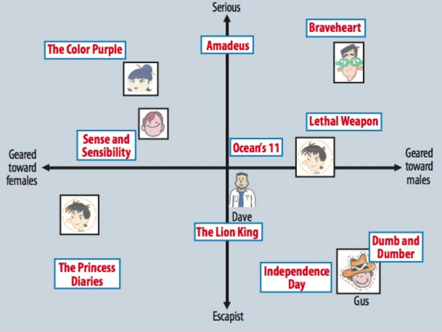


Figure 2. A simplified illustration of the latent factor approach, which characterizes both users and movies using two axes—male versus female and serious versus escapist.

Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," Computer, vol. 42, no. 8, pp. 30-37, August 2009

The Netflix Competition

- Problem: prediction of user ratings for films (collaborative filtering)
- Prize: \$1 Million to the first algorithm able to improve Netflix own algorithm results by at least 10%

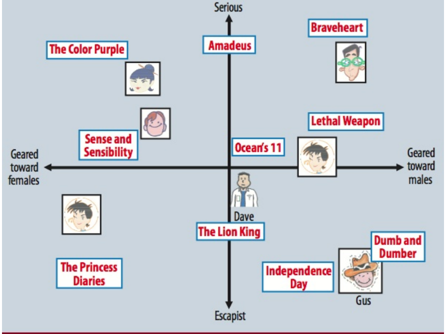


Figure 2. A simplified illustration of the latent factor approach, which characterizes both users and movies using two axes—male versus female and serious versus escapist.

Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," Computer, vol. 42, no. 8, pp. 30-37, August 2009

The Netflix Competition

- Problem: prediction of user ratings for films (collaborative filtering)
- Prize: $1 Million to the first algorithm able to improve Netflix own algorithm results by at least 10%
- Won on Sept/21/2009 by BellKor's Pragmatic Chaos

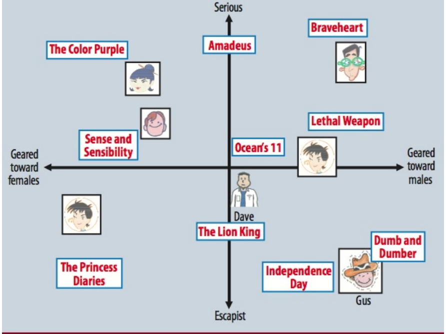


Figure 2. A simplified illustration of the latent factor approach, which characterizes both users and movies using two axes—male versus female and serious versus escapist.

Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," Computer, vol. 42, no. 8, pp. 30-37, August 2009

The Netflix Competition

- Problem: prediction of user ratings for films (collaborative filtering)
- Prize: $1 Million to the first algorithm able to improve Netflix own algorithm results by at least 10%
- Won on Sept/21/2009 by BellKor's Pragmatic Chaos
- Their approach used Non-negative Matrix Factorization (NMF) to build a latent-factor representation of users and movies

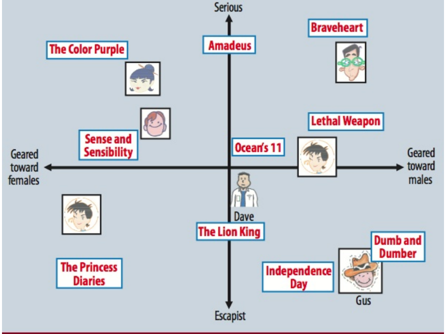


Figure 2. A simplified illustration of the latent factor approach, which characterizes both users and movies using two axes—male versus female and serious versus escapist.

Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," Computer, vol. 42, no. 8, pp. 30-37, August 2009

The Latent-Factor Model

$$
\begin{array}{cc}
 & \text{movies} \\
\text{users} & \begin{bmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & r_{nm} \end{bmatrix}
\end{array}
=
\begin{array}{cc}
 & \text{factors} \\
\text{users} & \begin{bmatrix} q_{11} & \cdots & q_{1f} \\ \vdots & \ddots & \vdots \\ q_{n1} & \cdots & q_{nf} \end{bmatrix}
\end{array}
\times
\begin{array}{cc}
 & \text{movies} \\
\text{factors} & \begin{bmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & \ddots & \vdots \\ p_{f1} & \cdots & p_{fm} \end{bmatrix}
\end{array}
$$

$$R \approx QP$$

$$Q, P \geq 0$$

$$
\begin{aligned}
n &\approx 5 \times 10^5 \\
m &\approx 1.7 \times 10^4 \\
|\{(i,j) | r_{ij} \neq 0\}| &\approx 10^8 \\
f &\leq 200
\end{aligned}
$$

Outline

Non-negative Matrix Factorization

- Problem: to find a factorization

$$X_{n\times m} = W_{n\times r}H_{r\times m}$$

- Optimization problem:

$$\begin{array}{ll} \min_{A,B} & ||X - WH||^2 \\ \text{s.t.} & W, H \geq 0 \end{array}$$

- $||\cdot||$ is the Frobenius norm
- It is a non-convex optimization problem
- Solution alternatives:
  - Gradient descendent methods
  - Multiplicative updating rules

Multiplicative Rules

- Optimization problem:

$$\begin{array}{cc} \min_{A,B} & ||X - WH||^2 \\ \text{s.t.} & W, H \geq 0 \end{array}$$

- Incremental optimization:

$$H_{a\mu} \leftarrow H_{a\mu} \frac{(W^T X)_{a\mu}}{(W^T W H)_{a\mu}}$$

$$W_{ia} \leftarrow W_{ia} \frac{(X H^T)_{\alpha\mu}}{(W H H^T)_{\alpha\mu}}$$

Divergence Optimization

- Optimization problem:

$$\begin{array}{ll} \min_{A,B} & D(X|WH) = \sum_{ij} \left(X_{ij} \log \frac{X_{ij}}{(WH)_{ij}} - X_{ij} + (WH)_{ij} \right) \\ \text{s.t.} & W, H \geq 0 \end{array}$$

- Multiplicative Rules:

$$H_{a\mu} \leftarrow H_{a\mu} \frac{\sum_i W_{ia} X_{i\mu}/(WH)_{i\mu}}{\sum_i W_{ia}}$$

$$W_{ia} \leftarrow W_{ia} \frac{\sum_\mu H_{a\mu} X_{i\mu}/(WH)_{i\mu}}{\sum_\mu H_{a\mu}}$$

Outline

PCA and SVD

- Problem:

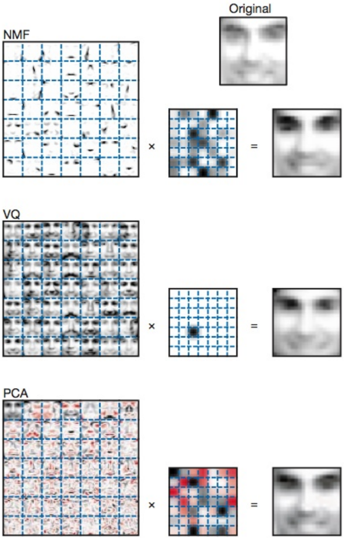
$$X_{n\times m} = W_{n\times r} H_{r\times m}$$

- Principal Component Analysis (PCA)

$$X = U\Sigma V$$

$$W = U\Sigma^{\frac{1}{2}}, H = \Sigma^{\frac{1}{2}} V$$

- PCA = SVD keeping the 'best' Eigenvectors
- Columns of U are orthonormal
- There is not restriction on sign.



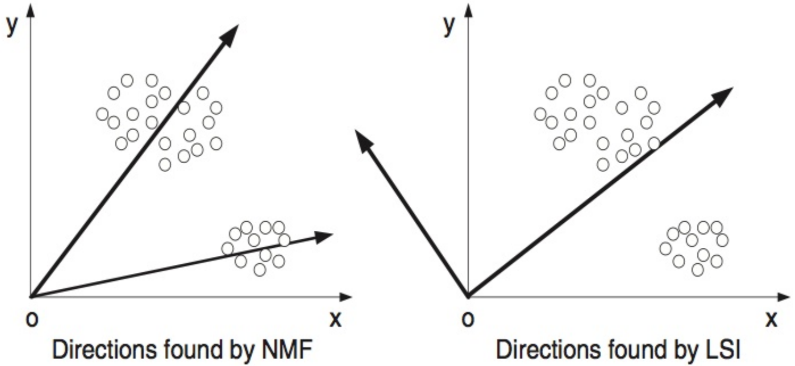
D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," Nature, vol. 401, no. 6755, pp. 788-791, October 1999

Latent Semantic Indexing (LSI)

$$\begin{array}{c} \\ \text{terms} \end{array} \underbrace{\overset{\text{documents}}{\begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}}}_{X} = \begin{array}{c} \\ \text{terms} \end{array} \underbrace{\overset{\text{factors}}{\begin{bmatrix} w_{11} & \cdots & w_{1r} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nr} \end{bmatrix}}}_{W} \times \begin{array}{c} \\ \text{factors} \end{array} \underbrace{\overset{\text{documents}}{\begin{bmatrix} h_{11} & \cdots & h_{1m} \\ \vdots & \ddots & \vdots \\ h_{r1} & \cdots & h_{rm} \end{bmatrix}}}_{H}$$

- Documents are represented by the frequency of keywords (terms)
- Uses SVD to find the factorization
- Factors = semantic concepts
- Columns of W are orthonormal

NMF vs LSI



W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," in SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval. New York, NY, USA: ACM, 2003, pp. 267-273

Outline

Outline

Semantic Space: Multimodal Latent Indexing (I)

- Objects are described by terms in a textual vocabulary and a visual vocabulary

Semantic Space: Multimodal Latent Indexing (I)

- Objects are described by terms in a textual vocabulary and a visual vocabulary
- Objects are mapped to a latent (semantic) space where objects are represented by a set of latent factors

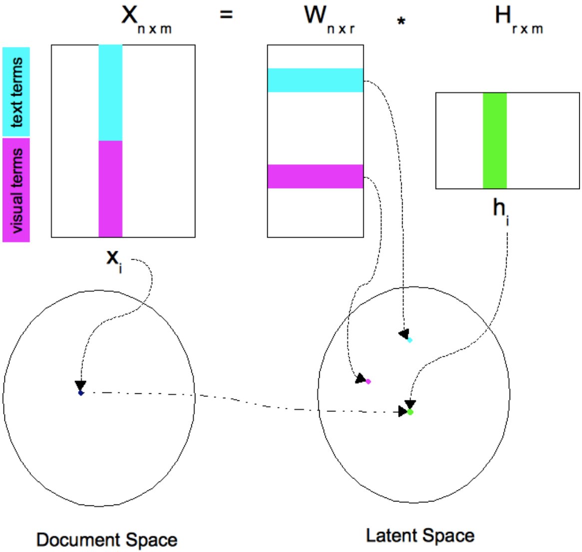
Semantic Space: Multimodal Latent Indexing (I)

- Objects are described by terms in a textual vocabulary and a visual vocabulary
- Objects are mapped to a latent (semantic) space where objects are represented by a set of latent factors
- NMF is used to build the latent representation

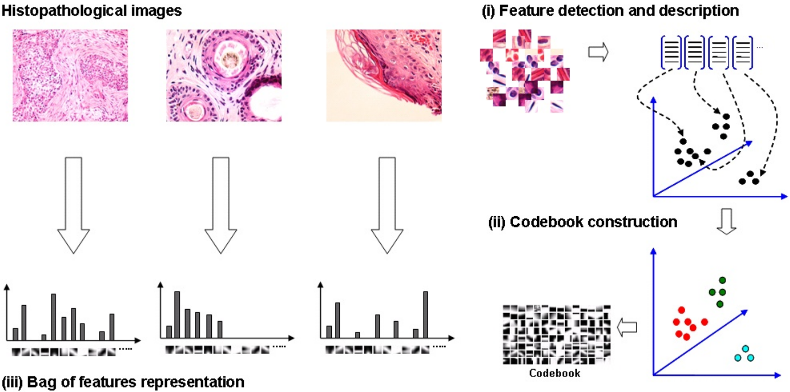
Semantic Space: Multimodal Latent Indexing (I)

- Objects are described by terms in a textual vocabulary and a visual vocabulary
- Objects are mapped to a latent (semantic) space where objects are represented by a set of latent factors
- NMF is used to build the latent representation
- Three main tasks:
  - Multimodal clustering
  - Automatic image annotation
  - Image retrieval

Semantic Space: Multimodal Latent Indexing (II)

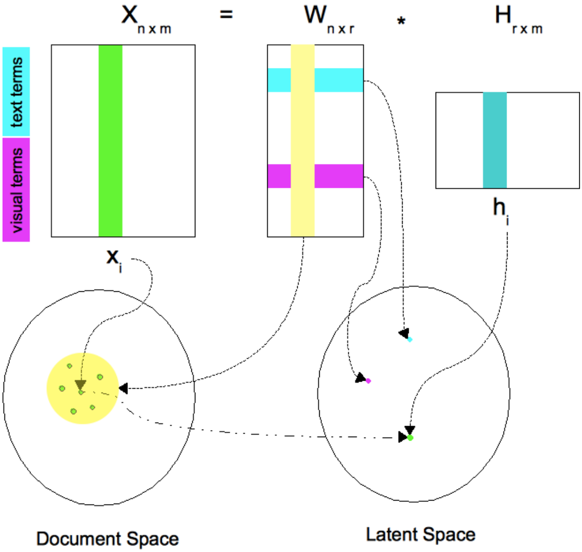


Bag-of-Features Image Representation

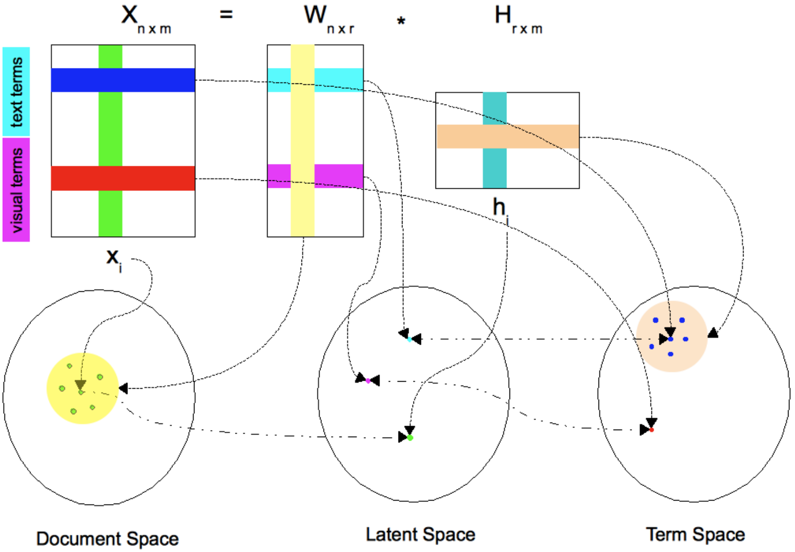


Outline

Multimodal Clustering



Dual Multimodal Clustering



Does It Work?

- Clustering Corel data set: 1000 images, 25 categories
- Input matrix: 2500 vectors of dimension 1000
- Clustering performance comparison: K-means and two NMF variants:

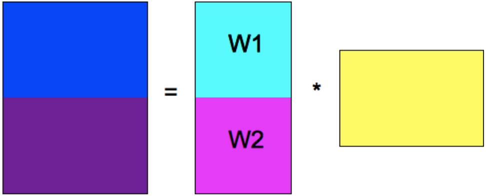
	K-means	NMF (Frobenius norm)	NMF (KL divergence)
Accuracy	0.2288	0.2768	0.2905

*results are average of 10 runs

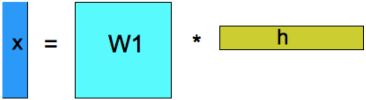
Outline

Image Annotation

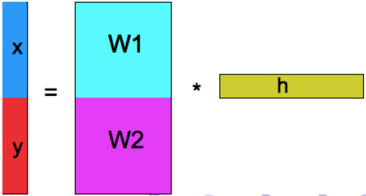
1. Apply NMF to training data



2. Find latent representation *h* of a visual vector *x*, *x = W1 * h*



3. Multiply *h* by *W* to get the multimodal vector *[x,y]*



Does it Work?

- pLSA has shown good performance for image annotation

TABLE 3
mAP Values (in Percent) for the Six Methods
When Combinations of HS and SIFT Features Are Used

	Blobs	HS	SIFT	HS+SIFT
propagation	7.8 (0.7)	9.0 (0.2)	9.4 (1.0)	13.1 (0.5)
CMRM [15]	11.5 (1.1)	10.7 (1.1)	7.9 (0.5)	13.4 (1.0)
SVD-COS [26]	12.9 (1.1)	12.9 (0.8)	10.7 (0.7)	16.6 (1.1)
PLSA-MIXED	5.8 (0.8)	10.2 (0.8)	7.5 (0.6)	11.9 (1.3)
PLSA-FEATURES	8.2 (0.7)	11.2 (1.0)	10.1 (0.8)	14.0 (1.3)
PLSA-WORDS	11.0 (0.9)	13.3 (1.0)	11.8 (1.1)	19.1 (1.2)

- NMF with KL divergence was shown to be equivalent to pLSA

W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," in SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval. New York, NY, USA: ACM, 2003, pp. 267-273

Outline

Image Retrieval

- Scenario: visual retrieval based on a visual query

Image Retrieval

- Scenario: visual retrieval based on a visual query
- Images are represented in a latent space using NMF

Image Retrieval

- Scenario: visual retrieval based on a visual query
- Images are represented in a latent space using NMF
- Some images in the database may not have text content associated

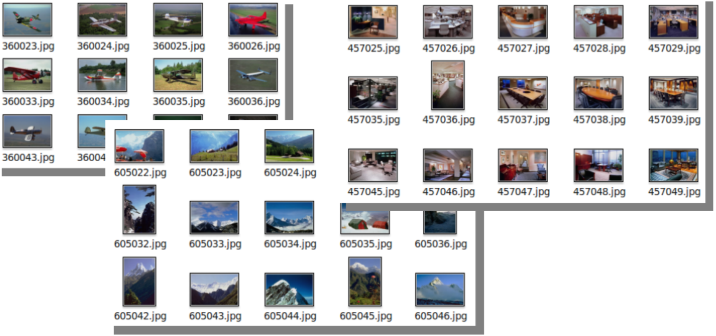
Image Retrieval

- Scenario: visual retrieval based on a visual query
- Images are represented in a latent space using NMF
- Some images in the database may not have text content associated
- The image query is projected the latent space

Image Retrieval

- Scenario: visual retrieval based on a visual query
- Images are represented in a latent space using NMF
- Some images in the database may not have text content associated
- The image query is projected the latent space
- Image are retrieved according to their latent space similarity

Experimental Evaluation



Corel Images

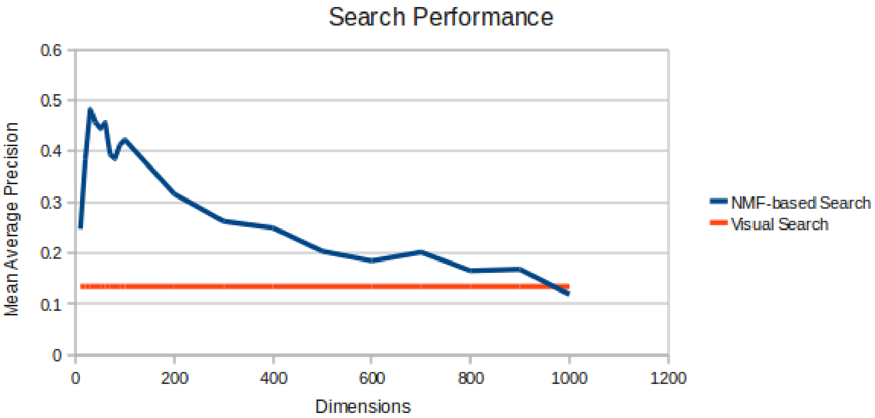
- A subset of 2,500 images extracted from the Corel Database.
- 25 categories with 100 images each.

Bag of Features

Corel Images

- Image content representation using parts of images
- Blocks are extracted from each image and the SIFT descriptor is computed
- A dictionary of visual patterns is built using k-means
- A histogram counting the occurrence of each codeblock is constructed for each image

Results



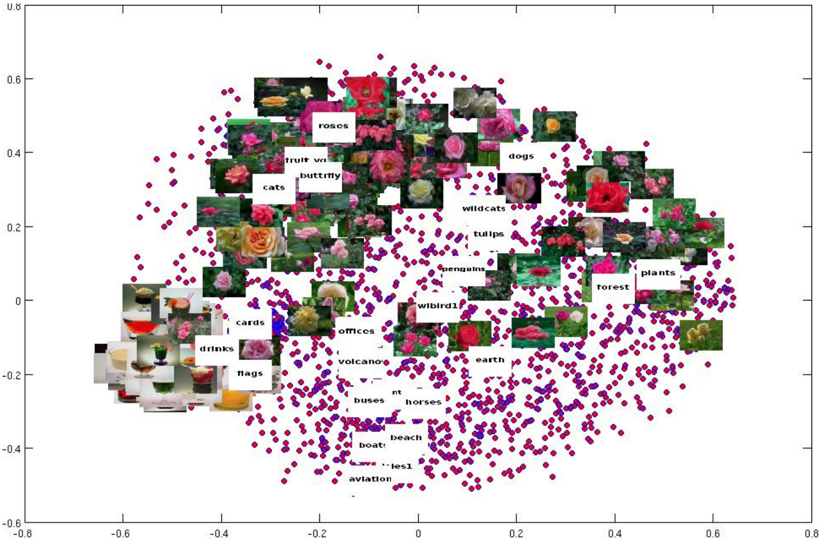
This approach outperforms direct image matching by almost 4X

Results

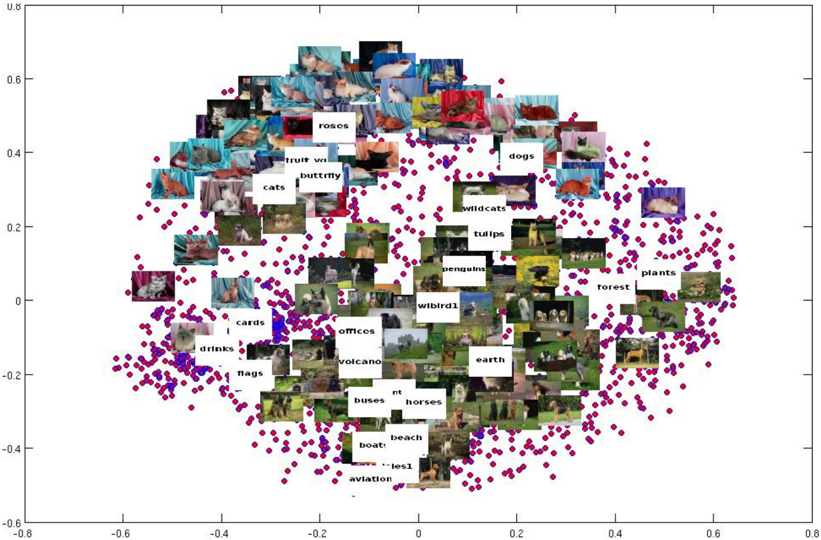
Some specific performance measures

Measure	NMF Search P1	Direct Search
P1	0.4400	0.5280
P20	0.4508	0.3820
P50	0.4474	0.3160
R10	0.0471	0.0446
R20	0.0949	0.0804
R50	0.2355	0.1663
MAP	0.4825	0.1342

Semantic Space Visualization (I)



Semantic Space Visualization (II)



Ongoing Work

- Bigger data sets: ImageClefMed, Corel5000, Flickr
- Different image low-level features
- Kernel NMF
- Incremental NMF

Thanks!

fagonzalezo@unal.edu.co

http://dis.unal.edu.co/~fgonza