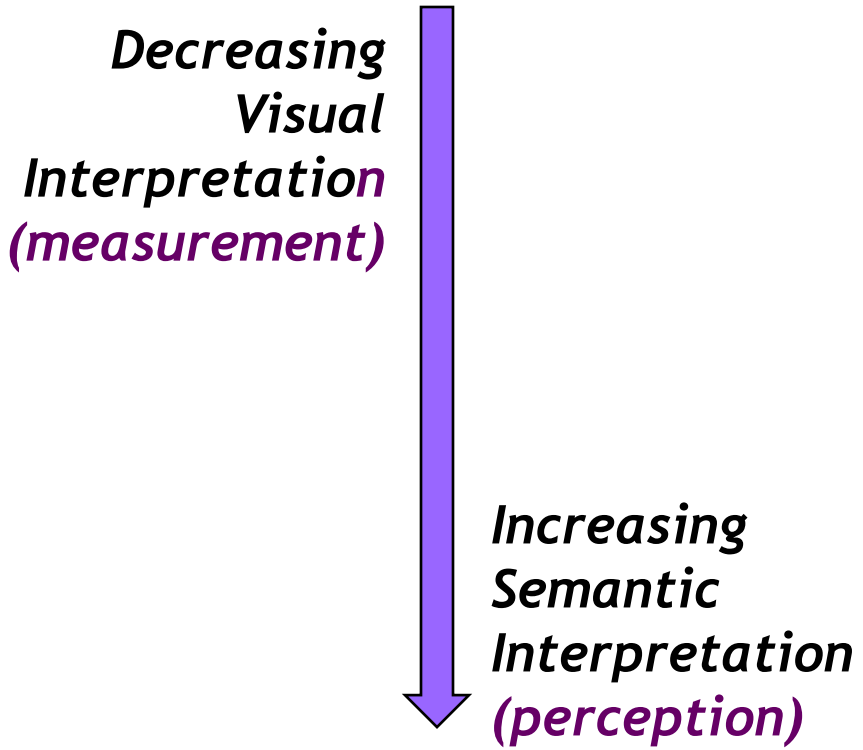
Cross-Modal Multimedia Retrieval

Nikhil Rasiwasia
Yahoo! Labs Bangalore

Visual Recognition

"Making sense of images"

- Finding Duplicates
- Image Retrieval
- Object Retrieval
- Image Classification
- Image Annotation
- Cross-Modal Retrieval



Visual Recognition

- **Finding Duplicates**
  - Find me the same image
  - Can be solved using md5 hash string matching.
  - So, md5 is the **“image signature”**
  - String matching, the **“similarity metric”**
  - Practically solved.

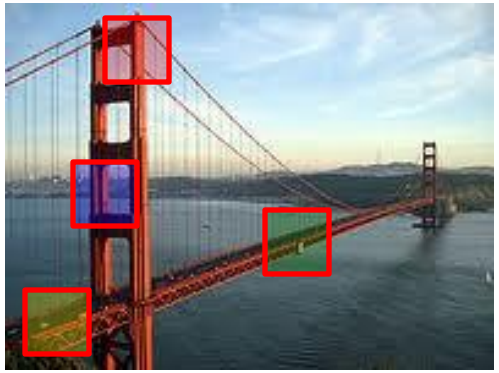
Visual Recognition

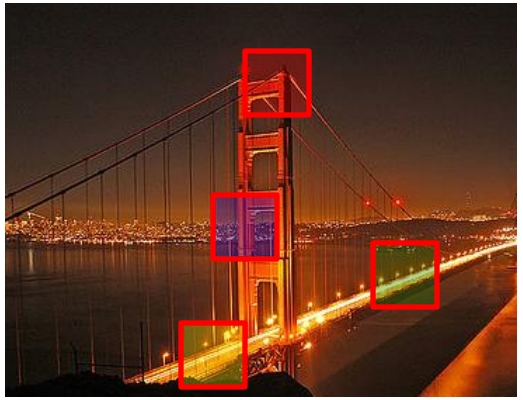
- **Image Retrieval**
  - Find me the similar images
  - Design **“image signatures”** and **“similarity metrics”**
  - “**invariant”** to scene variations
  - invariance is the key

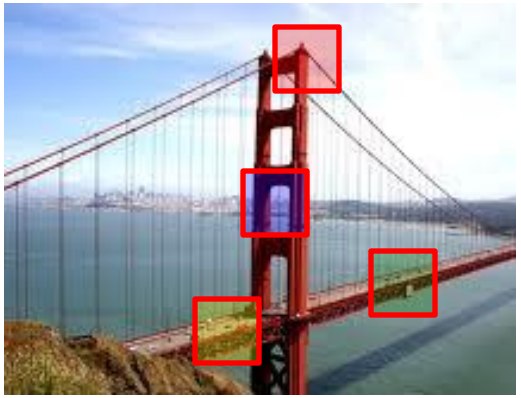
  - Global visual interpretation.

Visual Recognition

- **Object Retrieval**
  - Find me objects similar to this
  - Image signatures should be invariant to **“local”** scene variations
  - Local visual interpretation.

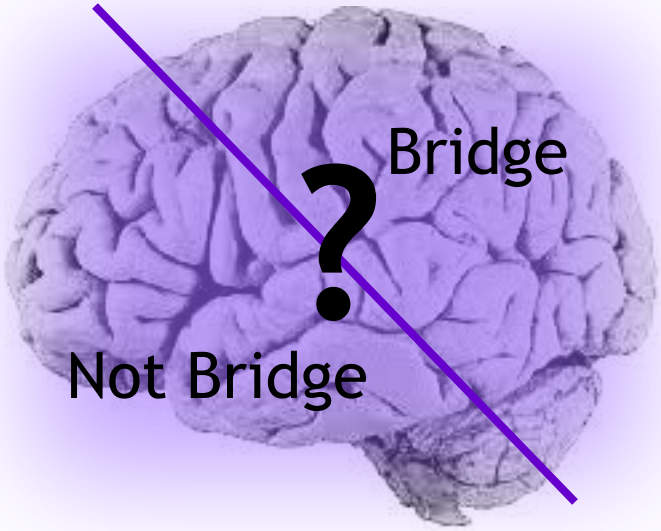






Visual Recognition

- **Image Classification**
  - Find me the category of the image/objects inside the image
  - Given
    - a **taxonomy** / a set of **classes**
    - a **labeled dataset** of images
  - Learn '**classifiers**'
  - Global/Local semantic interpretation



Visual Recognition

- **Image Annotation**
  - Find me tags for the image
  - Multi-label prediction task
  - Approach1: Classification
    - Learn classifier for each tag
    - Tag correlations are not used
  - Other Approaches
    - Nearest Neighbor based tag transfer

Bridge, Sky, Vacation, Golden Gate, San Francisco, etc.

  - Multiple semantic interpretation.

Visual Recognition

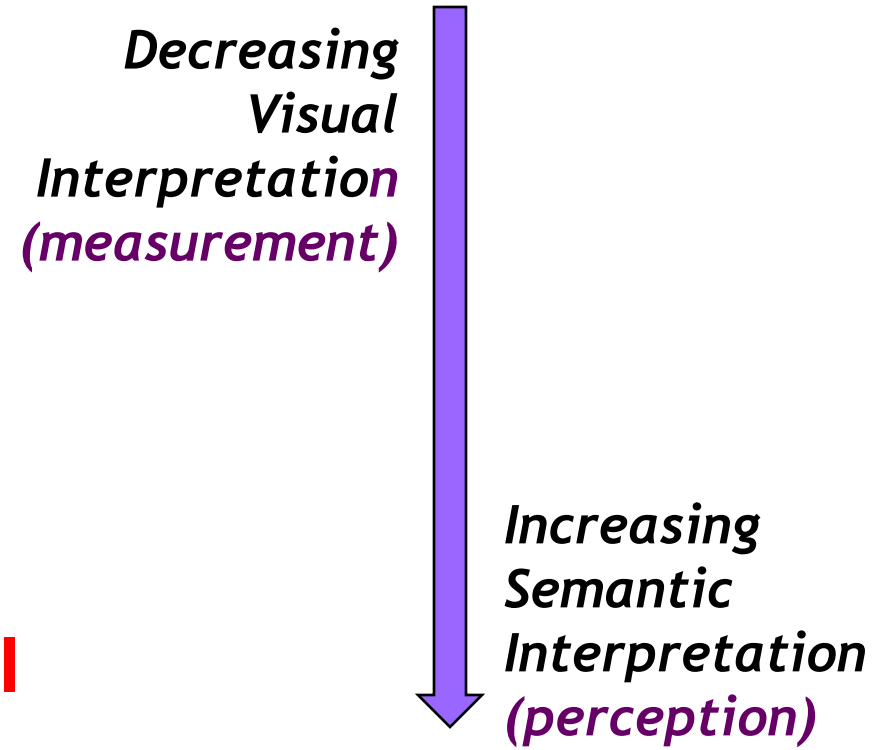
- **Cross-modal Retrieval**
  - Find me text articles that best describes the image.
  - Goes beyond multi-label and one step short of image description.
  - Complex semantic interpretation.

The Golden Gate Bridge is a suspension bridge spanning the Golden Gate, the opening of the San Francisco Bay into the Pacific Ocean. As part of both U.S. Route 101 and California State Route 1, the structure links the city of San Francisco, on the northern tip of the San Francisco Peninsula, to Marin County. It is one of the most internationally recognized symbols of San Francisco, California, and the United States. It has been declared one of the modern Wonders of the World by the American Society of Civil Engineers. The Frommers travel guide considers the Golden Gate Bridge "possibly the most beautiful, certainly the most photographed, bridge in the world".

Visual Recognition

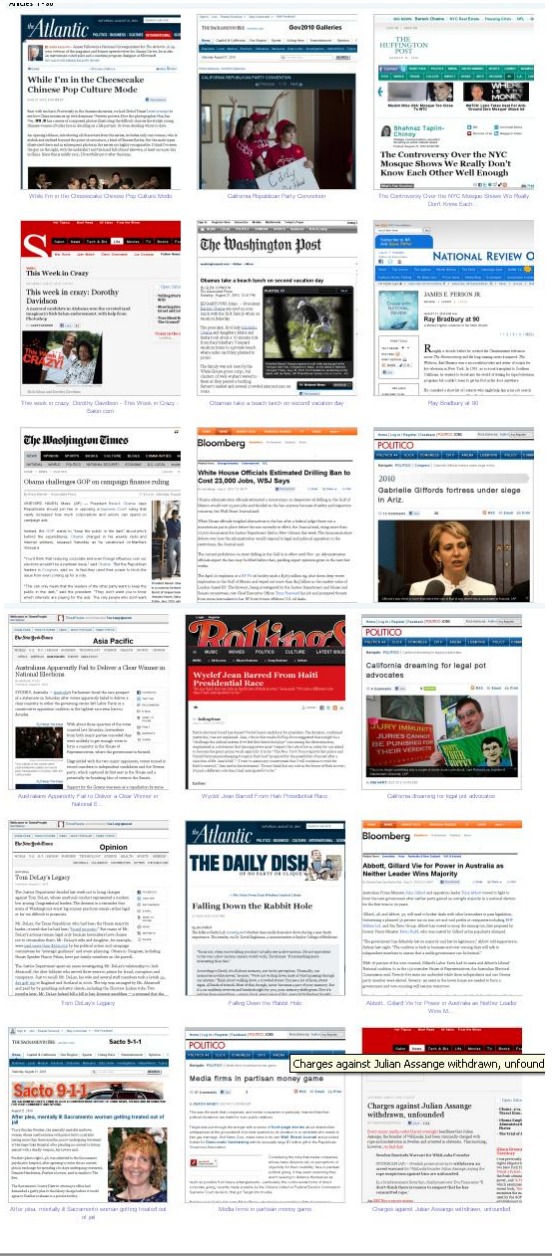
"Making sense of Images"

- Finding Duplicates
- Image Retrieval
- Object Retrieval
- Image Classification
- Image Annotation
- **Cross-Modal Retrieval**



Motivation

- Massive explosion of “content” on the web.
- Content rich in multiple modalities — Text, Images, Videos, Music etc.
- There is a need for retrieval systems that are transparent to modalities.

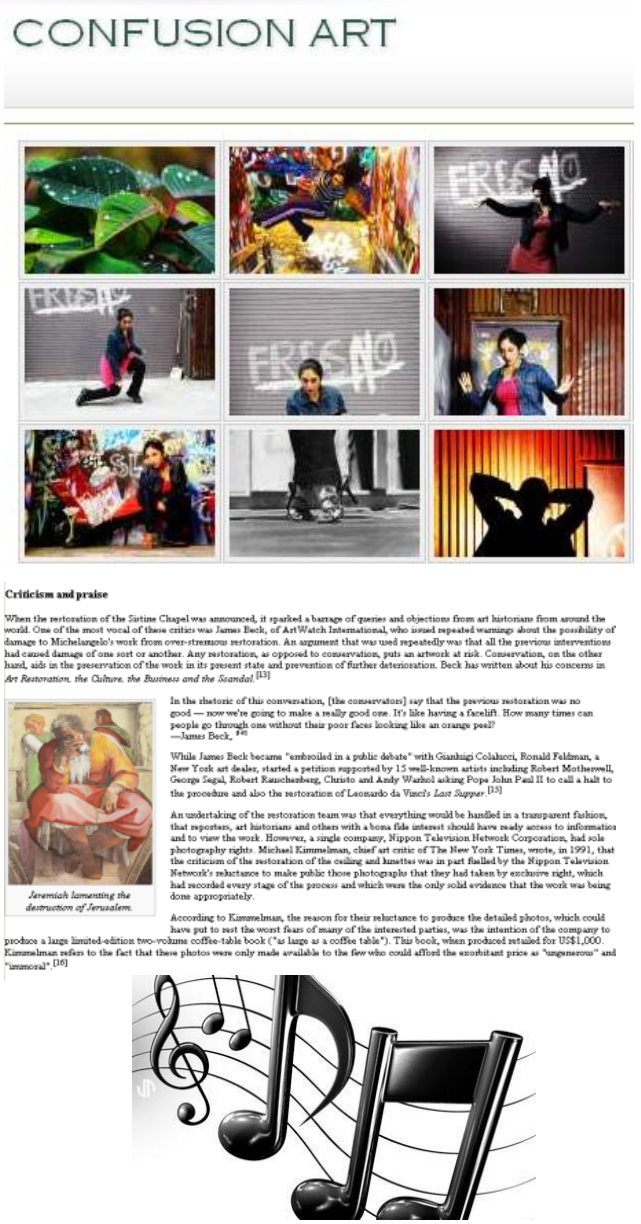


“Cross Modal Retrieval System”

Retrieval system that operates across multiple modalities

Motivation

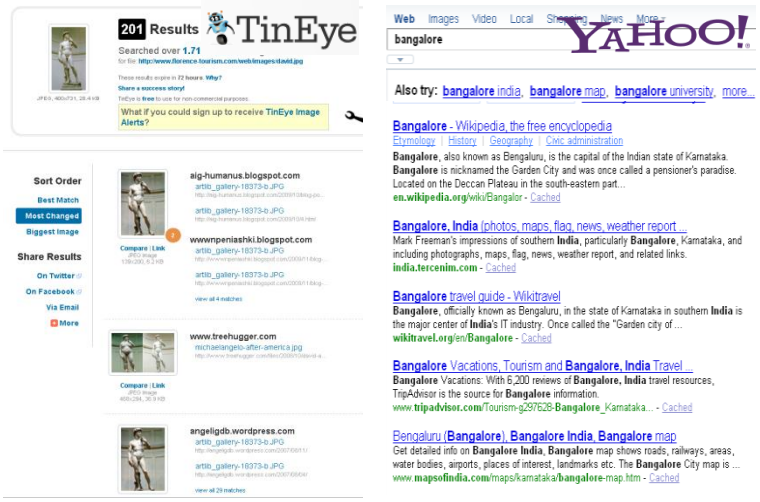
- Cross modal text query, eg. retrieval of images from photoblogs using textual query.
- Finding images to go along with a text article
- Finding music to enhance videos, slide shows.
- Image positioning.
- Text summarization based on images … and more



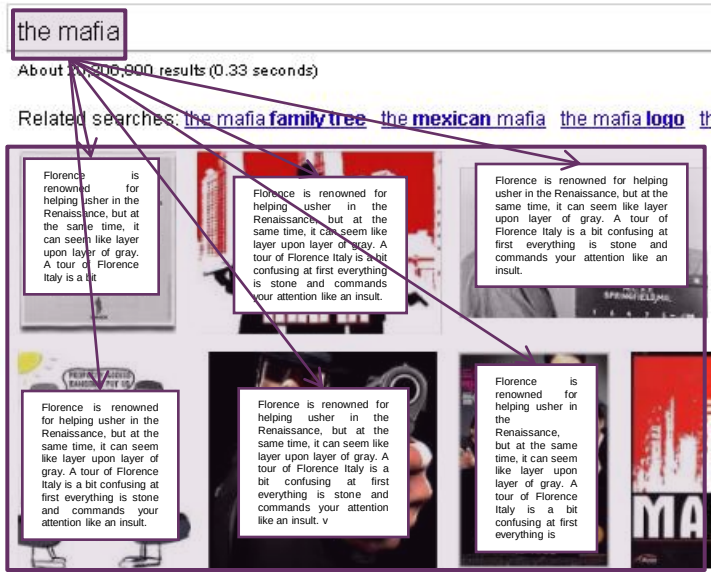
Current Retrieval Systems

- Current retrieval systems are predominantly uni-modal.
  - The query and retrieved results are from the same modality

Text	→	Text
Images	→	Images
Music	→	Music
Videos	→	Videos



- Is current Image search cross-modal retrieval?
  - **No,** text is matched to text metadata for the image
  - The operation would fail, in absence of text modality for the retrieval set.



Text → Text - - - Images

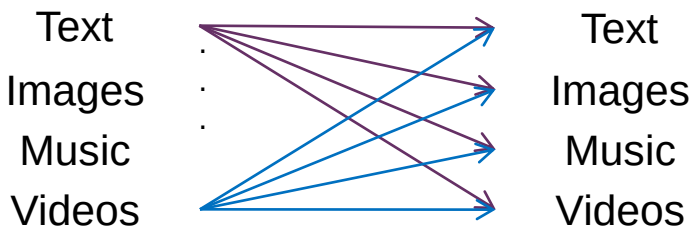
Current Retrieval Systems

- Several multi-modal systems have been proposed [TRECVID, ImageCLEF, Iria'09, Wang'09, Escalante'08, Pham'07, Snoek'05, Westerveld'02, etc.]
  - Given a query consisting of multiple modalities, retrieve examples containing the same multiple modalities.
  - Eg. Combining the modalities into a single modality, combining the outputs of multiple uni-modal systems.

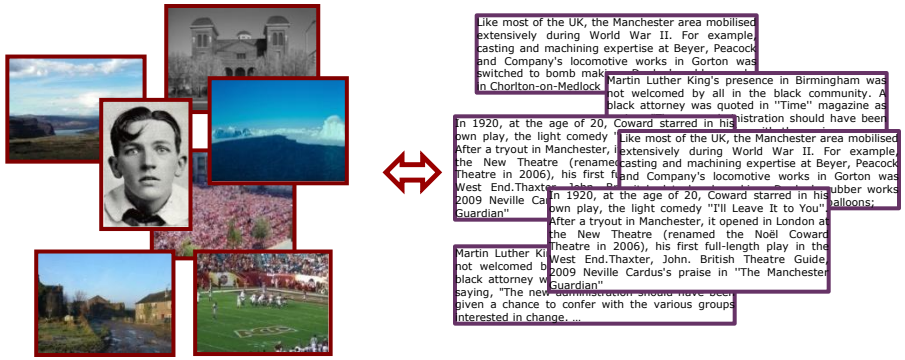
- Annotations systems [TRECVID, ImageCLEF, Carneiro'07, Feng'04, Lavrenko'03, Barnard'03, etc]
  - Given a query from a modality (say image), assign text labels.
  - Are **true cross-modal systems**.
  - However, text modality is constrained to a few keywords.

Cross Modal Retrieval

- Given query from modality A, retrieve results from modality B.
  - The query and retrieved items are not required to share a common modality.

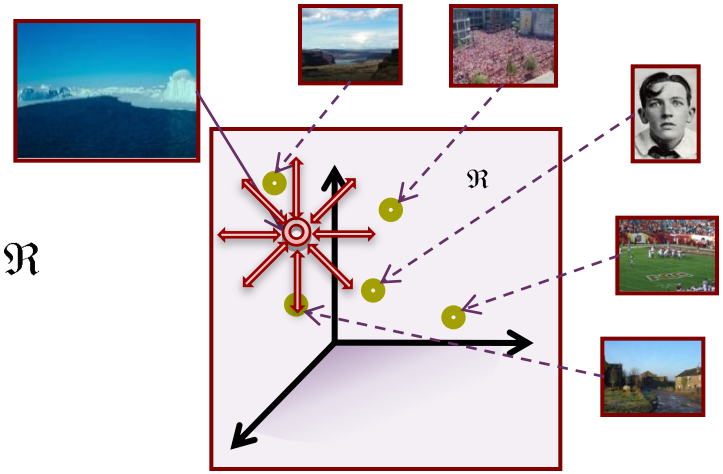


- In this tutorial – we restrict to text and image modalities
  - Although similar ideas can be applied to other modalities.
- Thus,
  - the retrieval of text in response to a query image.
  - And, the retrieval of images in response to a query text.

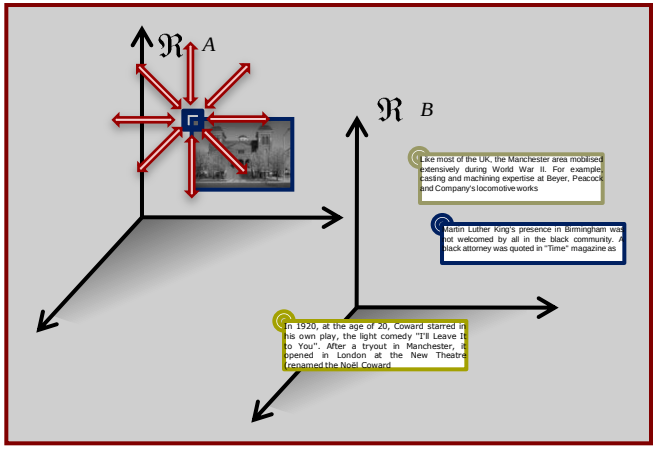


Design of Retrieval Systems

- Uni-modal Retrieval System
  - Design a feature space ($\Re$) for given modality
  - Map the query and retrieval set onto $\Re$
  - Using a suitable similarity function to rank the retrieval set.

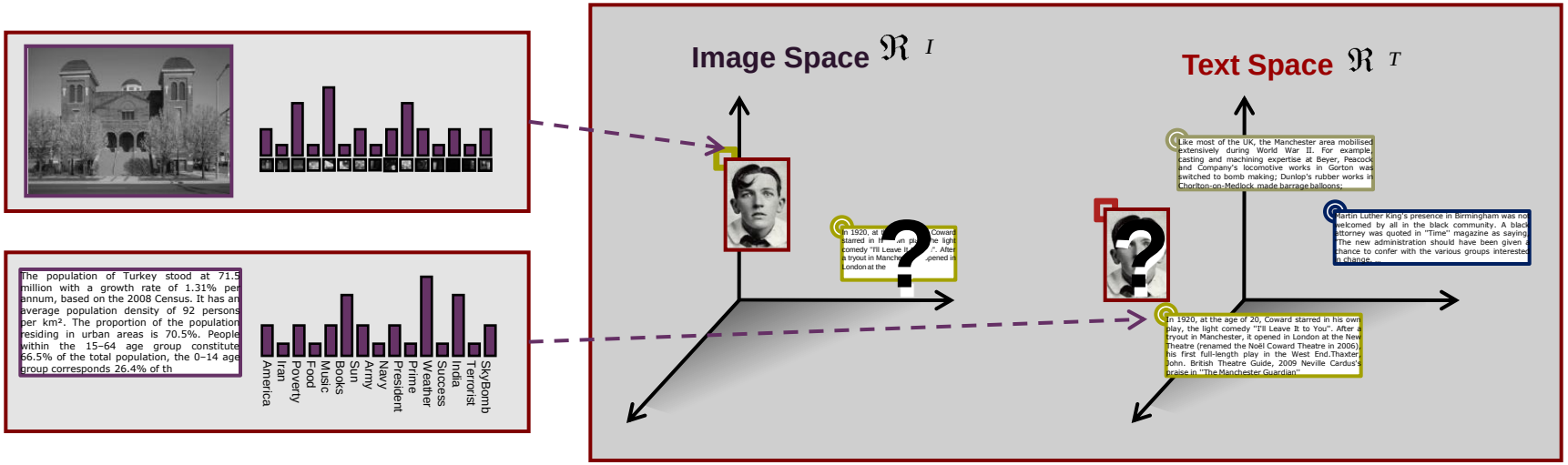


- Can this be applied to Cross Modal Retrieval?
  - Design feature spaces($\Re_A$, $\Re_B$) for two modalities.
  - Map query onto $\Re_A$ and the retrieval set onto $\Re_B$
  - But, what similarity function to use for ranking?



The problem.

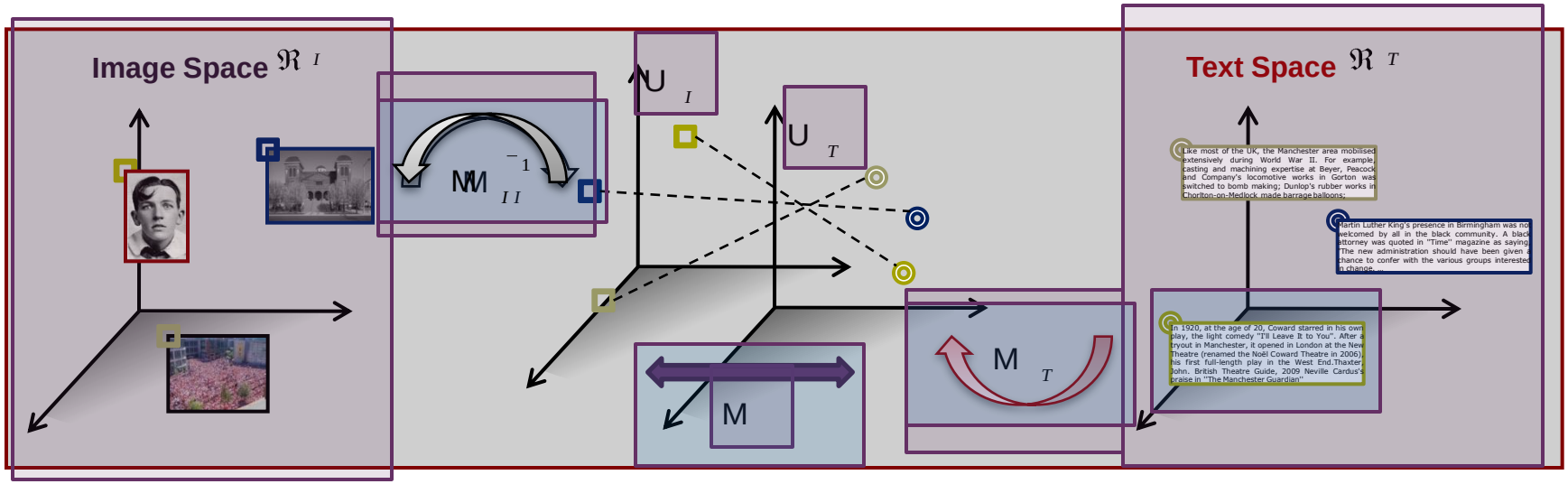
- No natural correspondence between representations of different modalities.
- For example, we use Bag-of-words representation for both images and text
  - Images: vectors over visual textures ($\Re^I$)
  - Text: vectors of word counts ($\Re^T$)



- How do we compute similarity?

An Idea

- **Learn mappings** (M_T, M_I) that maps different modalities into intermediate spaces (U_T, U_I) that have a natural and invertible correspondence (M)



- Given a text query T_q in $\Re^T$ the cross-modal retrieval reduces to find the nearest neighbor of: $M_I^{-1} \circ M \circ M_T (T_q)$
- Similarly for image query: $M_T^{-1} \circ M^{-1} \circ M_I (I_q)$
- ***The task now is to design these mappings.***

The Fundamental Hypotheses

- We explore two fundamental hypotheses

1. Correlation Matching (CM) Hypothesis: The problem is that there is no correlation between the representations of different modalities.

   Can be tested by designing intermediate representations that maximizes correlations between modalities.

2. Semantic Matching (SM) Hypothesis: The problem is that the representation lacks common semantics.

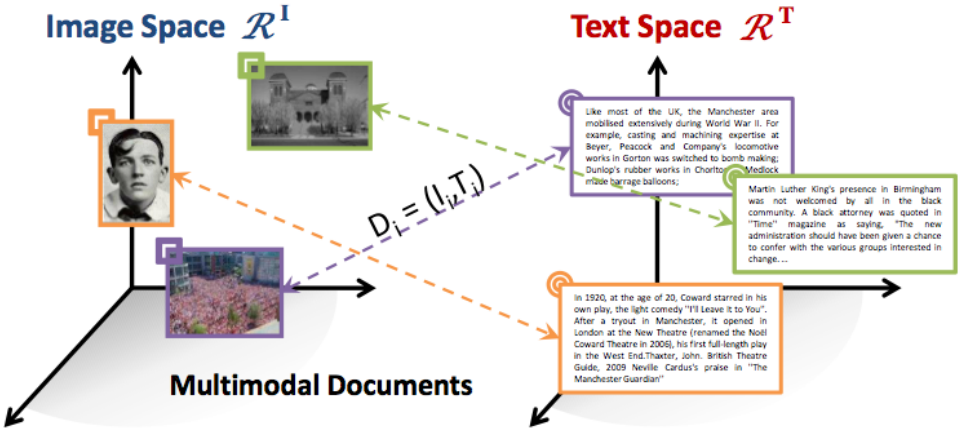
   Can be tested by designing a shared semantic representation for all modalities.

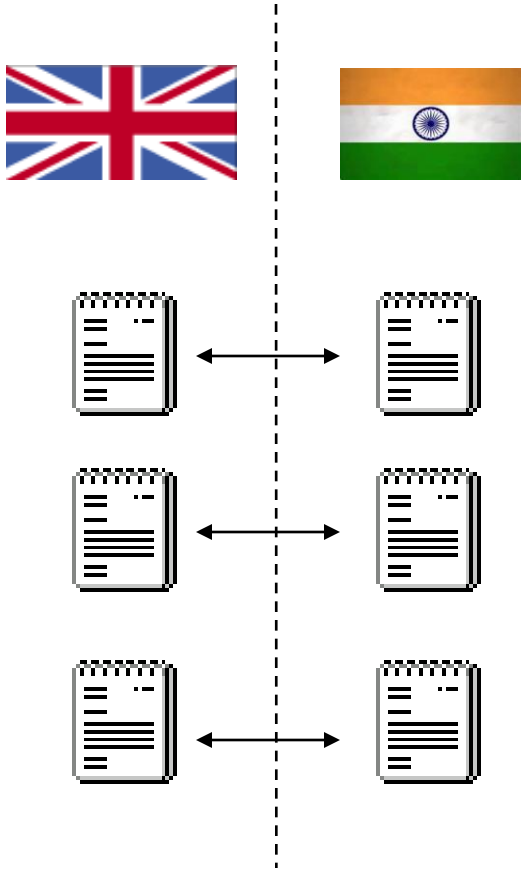
Correlation Matching (CM)

- Learn subspaces that maximize correlation between two modalities
- We use **Canonical Correlation Analysis** (CCA) to obtain mappings that maximize correlation.
- CCA was developed first by H. Hotelling.
  - H. Hotelling. Relations between two sets of variates. *Biometrika*, 28:321-377, 1936.
  - Commonly used in economics, medical studies, bioinformatics and other areas.

Canonical Correlation Analysis

- **Measures the linear relationship between two views of the same data**
  - Eg, Text documents written in different languages,
  - Eg, Images with attached text
  - Eg, Protein sequence and corresponding gene expression



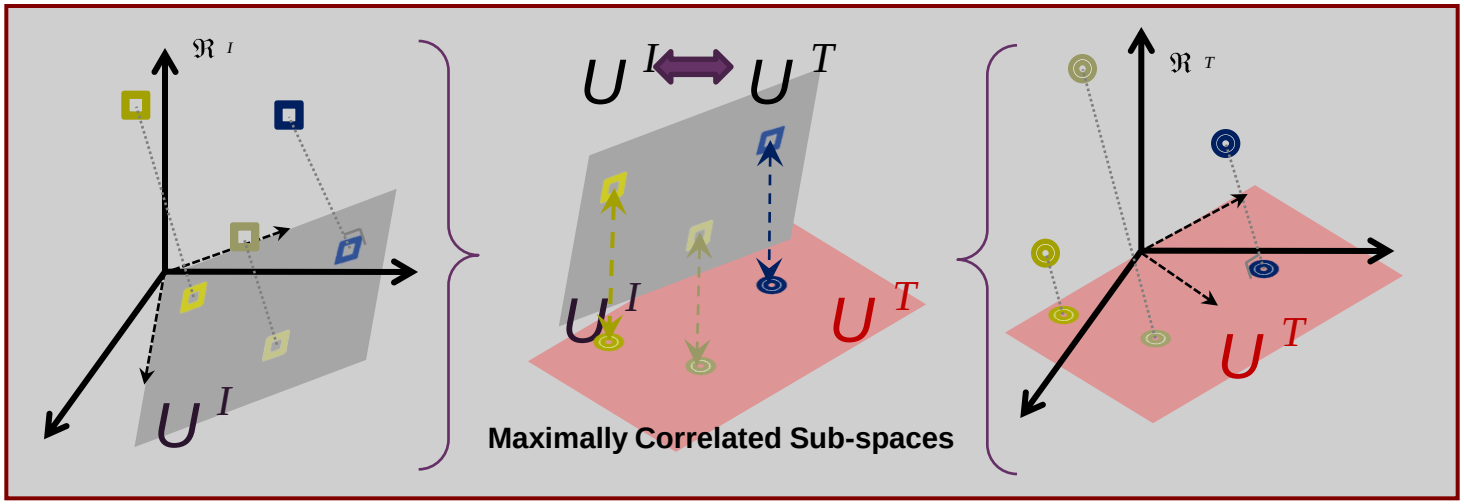


Canonical Correlation Analysis

- Finds two sets of basis vectors such that the **correlation between the projections** of the variables onto these basis vectors is maximized
  - Find pairs of features from both views with highest correlations
  - Example: words that co-appear in document and its translation

train	Sahasra Chakra Louh Path Gaamini
Mosquito	Gunjanhaari Manav Rakt Pipasu Jeev

Intuition

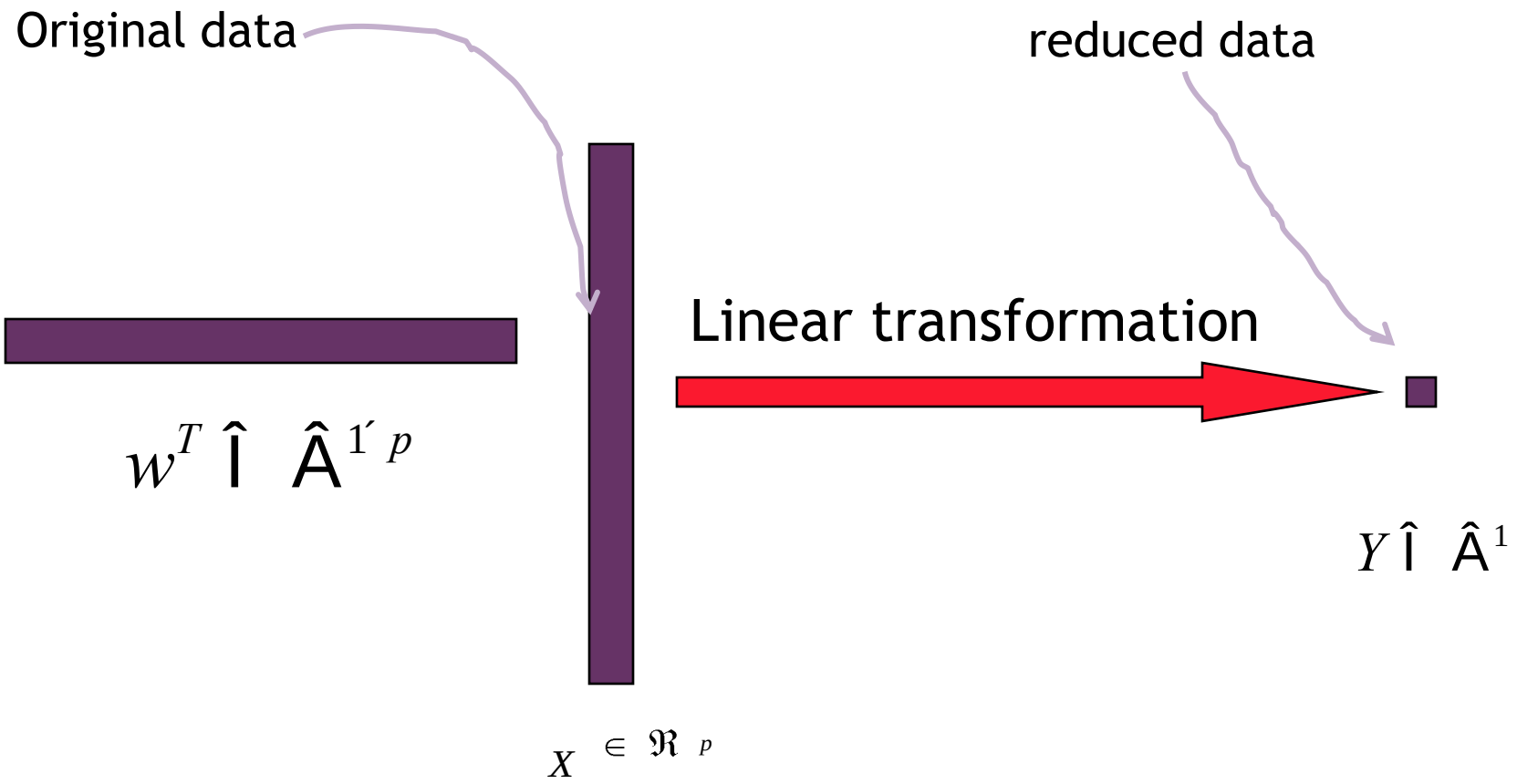


- Thus a technique for joint dimensionality reduction.

Overview of feature reduction

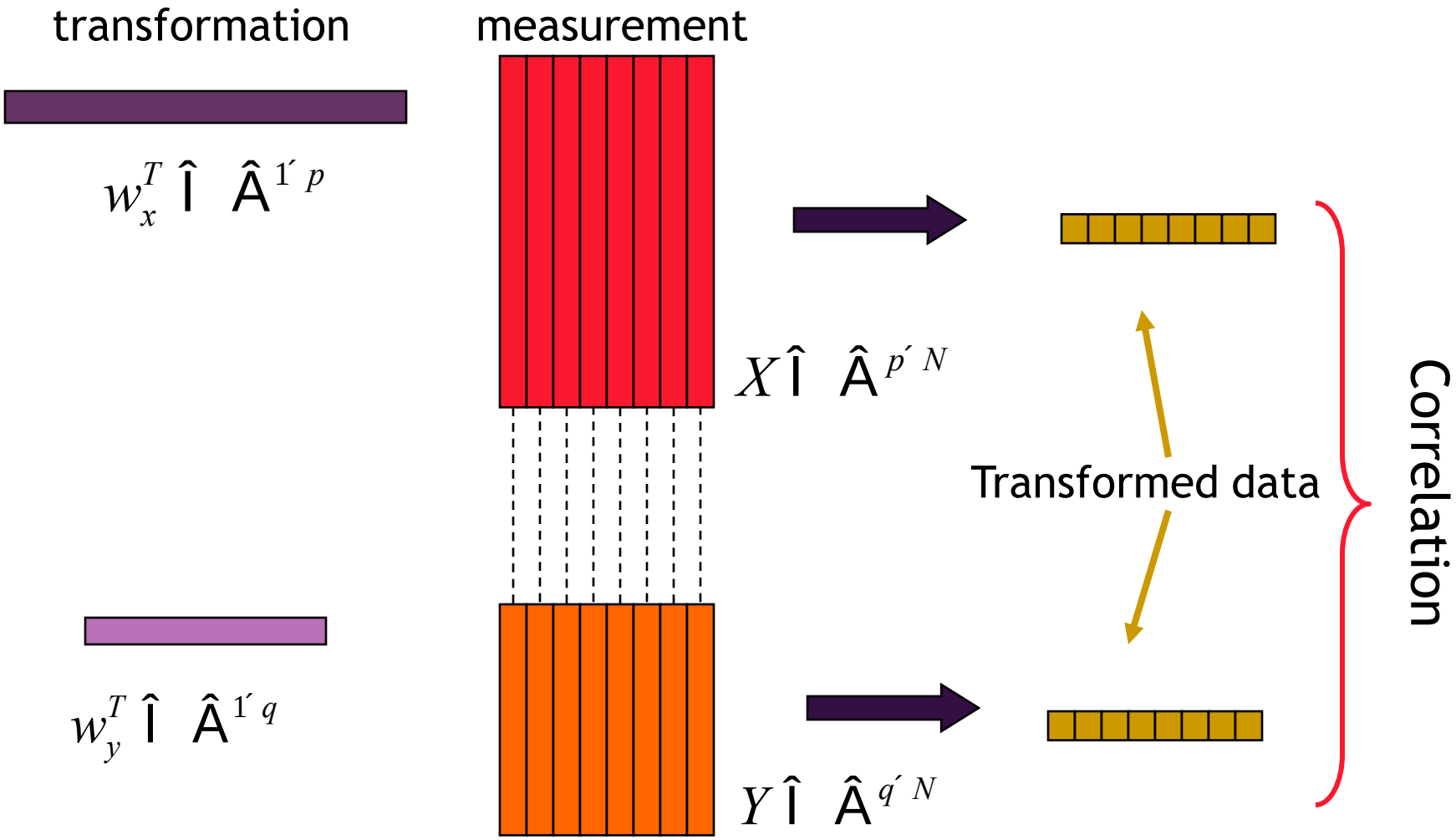
- Unsupervised for single dataset
  - Latent Semantic Indexing (LSI): truncated SVD
  - Principal Component Analysis (PCA)
- Supervised for single dataset
  - Linear Discriminant Analysis (LDA)
- Unsupervised for two dataset with pairwise correspondence
  - Canonical Correlation Analysis (CCA)

Overview of feature reduction



$$w \in \Re^{p \times 1} : X \to Y = w^T X \in \Re^1$$

Canonical Correlation Analysis (CCA)



Problem definition

- Find two sets of basis vectors, one for x and the other for y, such that the correlations between the *projections* of the variables onto these basis vectors are maximized.

Given $S = ((\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_n, \mathbf{y}_n)) \text{ of } (\mathbf{x}, \mathbf{y})$

Compute two basis vectors w_x and w_y :

$x \rightarrow < w_x, x >$ $\quad S_{x,\mathbf{w}_x} = (\langle \mathbf{w}_x, \mathbf{x}_1 \rangle, \ldots, \langle \mathbf{w}_x, \mathbf{x}_n \rangle)$

$y \rightarrow < w_y, y >$ $\quad S_{y,\mathbf{w}_y} = (\langle \mathbf{w}_y, \mathbf{y}_1 \rangle, \ldots, \langle \mathbf{w}_y, \mathbf{y}_n \rangle)$

Problem definition

- Compute the two basis vectors so that the correlations of the projections onto these vectors are maximized.

$$\rho = \max_{\mathbf{w}_x, \mathbf{w}_y} corr(S_x \mathbf{w}_x, S_y \mathbf{w}_y)$$

$$= \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\langle S_x \mathbf{w}_x, S_y \mathbf{w}_y \rangle}{\|S_x \mathbf{w}_x\| \|S_y \mathbf{w}_y\|}.$$

Algebraic derivation of CCA

Now observe that the covariance matrix of $(\mathbf{x}, \mathbf{y})$ is

$$C(\mathbf{x}, \mathbf{y}) = \hat{\mathbb{E}}\left[\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}'\right] = \begin{bmatrix}C_{\mathbf{xx}} & C_{\mathbf{xy}}\\ C_{\mathbf{yx}} & C_{\mathbf{yy}}\end{bmatrix} = C.$$

The optimization problem is equivalent to

$$\rho = \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\mathbf{w}_x' C_{\mathbf{xy}} \mathbf{w}_y}{\sqrt{\mathbf{w}_x' C_{\mathbf{xx}} \mathbf{w}_x \mathbf{w}_y' C_{\mathbf{yy}} \mathbf{w}_y}}.$$

where

$$C_{xy} = XY^T, \; C_{xx} = XX^T$$

$$C_{yx} = YX^T, \; C_{yy} = YY^T$$

Empirical covariance for images and text, and their cross covariance.

Derivation of CCA

Observe that the solution is not affected by a rescaling of w_x or w_y thus the optimization problem is equivalent to

$$\rho = \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\mathbf{w}_x' C_{xy} \mathbf{w}_y}{\sqrt{\mathbf{w}_x' C_{xx} \mathbf{w}_x \mathbf{w}_y' C_{yy} \mathbf{w}_y}}.$$

$$\begin{aligned} \max \quad & \mathbf{w}_x' C_{xy} \mathbf{w}_y \\ \text{s.t.} \quad & \mathbf{w}_x' C_{xx} \mathbf{w}_x = 1 \\ & \mathbf{w}_y' C_{yy} \mathbf{w}_y = 1. \end{aligned}$$

The corresponding Lagrangian is

$$L(\lambda, \mathbf{w}_x, \mathbf{w}_y) = \mathbf{w}_x' C_{xy} \mathbf{w}_y - \frac{\lambda_x}{2}(\mathbf{w}_x' C_{xx} \mathbf{w}_x - 1) - \frac{\lambda_y}{2}(\mathbf{w}_y' C_{yy} \mathbf{w}_y - 1).$$

Algebraic derivation of CCA

Taking derivatives in respect to $\mathbf{w}_x$ and $\mathbf{w}_y$, we obtain

$$\frac{\partial L}{\partial \mathbf{w}_x} = C_{xy}\mathbf{w}_y - \lambda_x C_{xx}\mathbf{w}_x = \mathbf{0}$$

$$\frac{\partial L}{\partial \mathbf{w}_y} = C_{yx}\mathbf{w}_x - \lambda_y C_{yy}\mathbf{w}_y = \mathbf{0}.$$

$C_{yx} = C_{xy}^T$ $\Rightarrow$ $\lambda_x = \lambda_y = \lambda$

$$C_{xy}C_{yy}^{-1}C_{yx}\mathbf{w}_x = \lambda^2 C_{xx}\mathbf{w}_x.$$

$$\mathbf{w}_y = \frac{C_{yy}^{-1}C_{yx}\mathbf{w}_x}{\lambda}$$

Nonlinear CCA using Kernels

Key: rewrite the CCA formulation in terms of inner products.

$$\rho = \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\mathbf{w}_x' C_{xy} \mathbf{w}_y}{\sqrt{\mathbf{w}_x' C_{xx} \mathbf{w}_x \mathbf{w}_y' C_{yy} \mathbf{w}_y}}.$$

$$C_{xx} = XX^T$$

$$C_{xy} = XY^T$$

$$w_x = X\alpha$$

$$w_y = Y\beta$$

$$\rho = \max_{\alpha, \beta} \frac{\alpha^T X^T XY^T Y \beta}{\sqrt{\alpha^T X^T XX^T X \alpha}\sqrt{\beta^T Y^T YY^T Y \beta}}$$

Only inner products Appear

Nonlinear CCA using Kernels

Recall that $$\rho = \max_{\alpha,\beta} \frac{\alpha^T X^T XY^T Y\beta}{\sqrt{\alpha^T X^T XX^T X\alpha}\sqrt{\beta^T Y^T YY^T Y\beta}}$$

Apply the following nonlinear transformation on x and y:

$$\phi_x : x_i \to \phi_x(x_i) \qquad \phi_y : y_i \to \phi_y(y_i)$$

Define the following two kernels:

$$K_x(x_1, x_2) = \phi_x(x_1) \bullet \phi_x(x_2)$$

$$K_y(y_1, y_2) = \phi_y(y_1) \bullet \phi_y(y_2)$$

$$\rho = \max_{\alpha,\beta} \frac{\alpha' K_x K_y \beta}{\sqrt{\alpha' K_x^2 \alpha \cdot \beta' K_y^2 \beta}}$$

Nonlinear CCA using Kernels

Define the Lagrangian as follows:

$$L(\lambda, \alpha, \beta) = \alpha' K_x K_y \beta - \frac{\lambda_\alpha}{2}(\alpha' K_x^2 \alpha - 1) - \frac{\lambda_\beta}{2}(\beta' K_y^2 \beta - 1)$$

Take the derivatives and set to 0:

$$\frac{\partial L}{\partial \alpha} = K_x K_y \beta - \lambda_\alpha K_x^2 \alpha = \mathbf{0}$$

$$\frac{\partial L}{\partial \beta} = K_y K_x \alpha - \lambda_\beta K_y^2 \beta = \mathbf{0}$$

$$K_x K_y K_y^{-1} K_x \alpha - \lambda^2 K_x K_x \alpha = 0$$

$$I\alpha = \lambda^2 \alpha$$

$$\beta = \frac{K_y^{-1} K_y^{-1} K_y K_x \alpha}{\lambda}$$

$$= \frac{K_y^{-1} K_x \alpha}{\lambda}.$$

Nonlinear CCA using Kernels

Two limitations: overfitting and singularity problem.
Solution: apply regularization technique to both x and y.

$$\rho = \max_{\alpha,\beta} \frac{\alpha' K_x K_y \beta}{\sqrt{(\alpha' K_x^2 \alpha + \kappa \|\mathbf{w}_x\|^2) \cdot (\beta' K_y^2 \beta + \kappa \|\mathbf{w}_y\|^2))}}$$

$$= \max_{\alpha,\beta} \frac{\alpha' K_x K_y \beta}{\sqrt{(\alpha' K_x^2 \alpha + \kappa \alpha' K_x \alpha) \cdot (\beta' K_y^2 \beta + \kappa \beta' K_y \beta)}}.$$

The solution is given by computing the following eigen-decomposition:

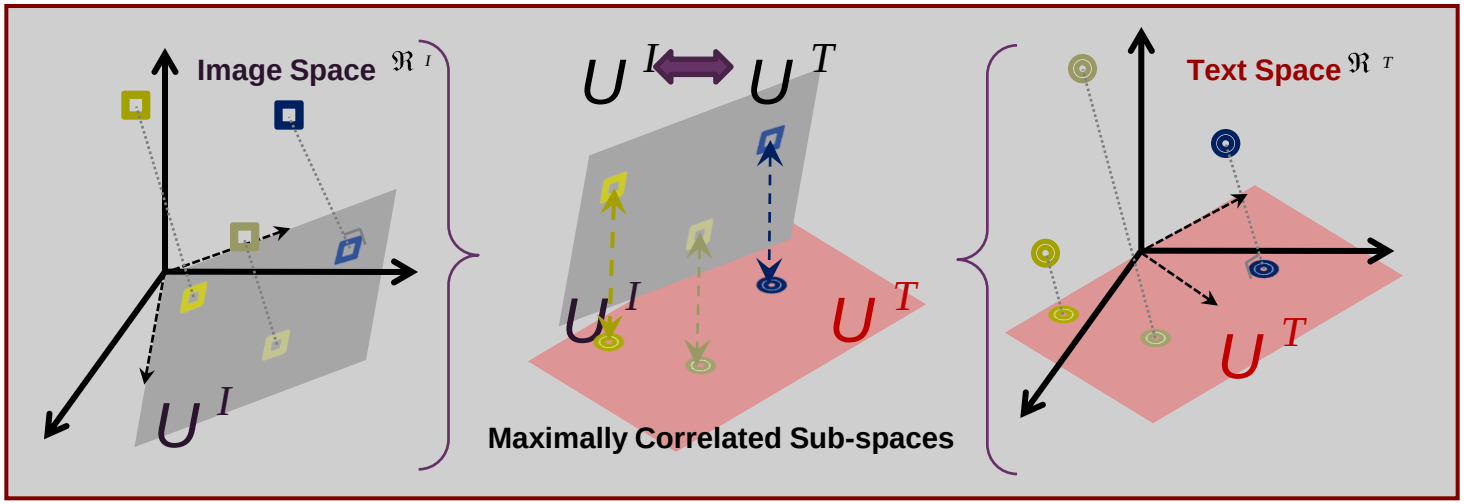
$$(K_x + \kappa I)^{-1} K_y (K_y + \kappa I)^{-1} K_x \alpha = \lambda^2 \alpha$$

Other Approaches

- Other approaches that maximize
  - Partial Least Squares
  - Bilinear Models
  - Canonical Covariance Analysis
  - Cross-modal Factor Analysis

- Will not cover.

Correlation Matching (CM)



The Fundamental Hypotheses

- We explore two fundamental hypotheses

1. Correlation Matching (CM) Hypothesis: The problem is that there is no correlation between the representations of different modalities.

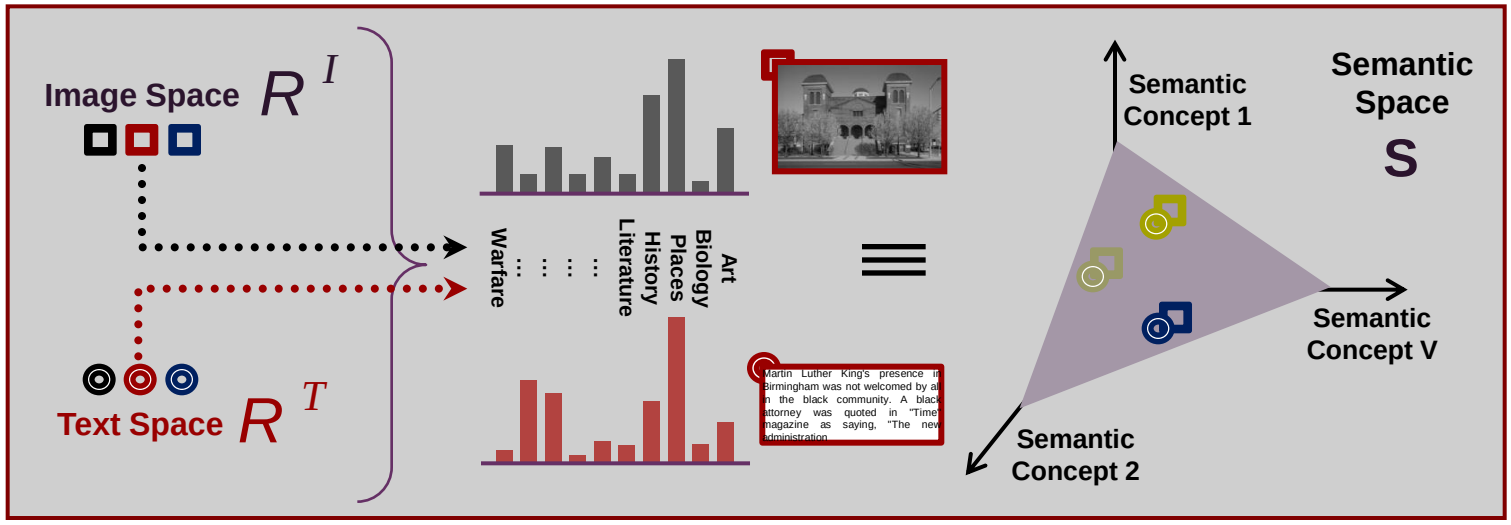
   Can be tested by designing intermediate representations that maximizes correlations between modalities.

2. Semantic Matching (SM) Hypothesis: The problem is that the representation lacks common semantics.

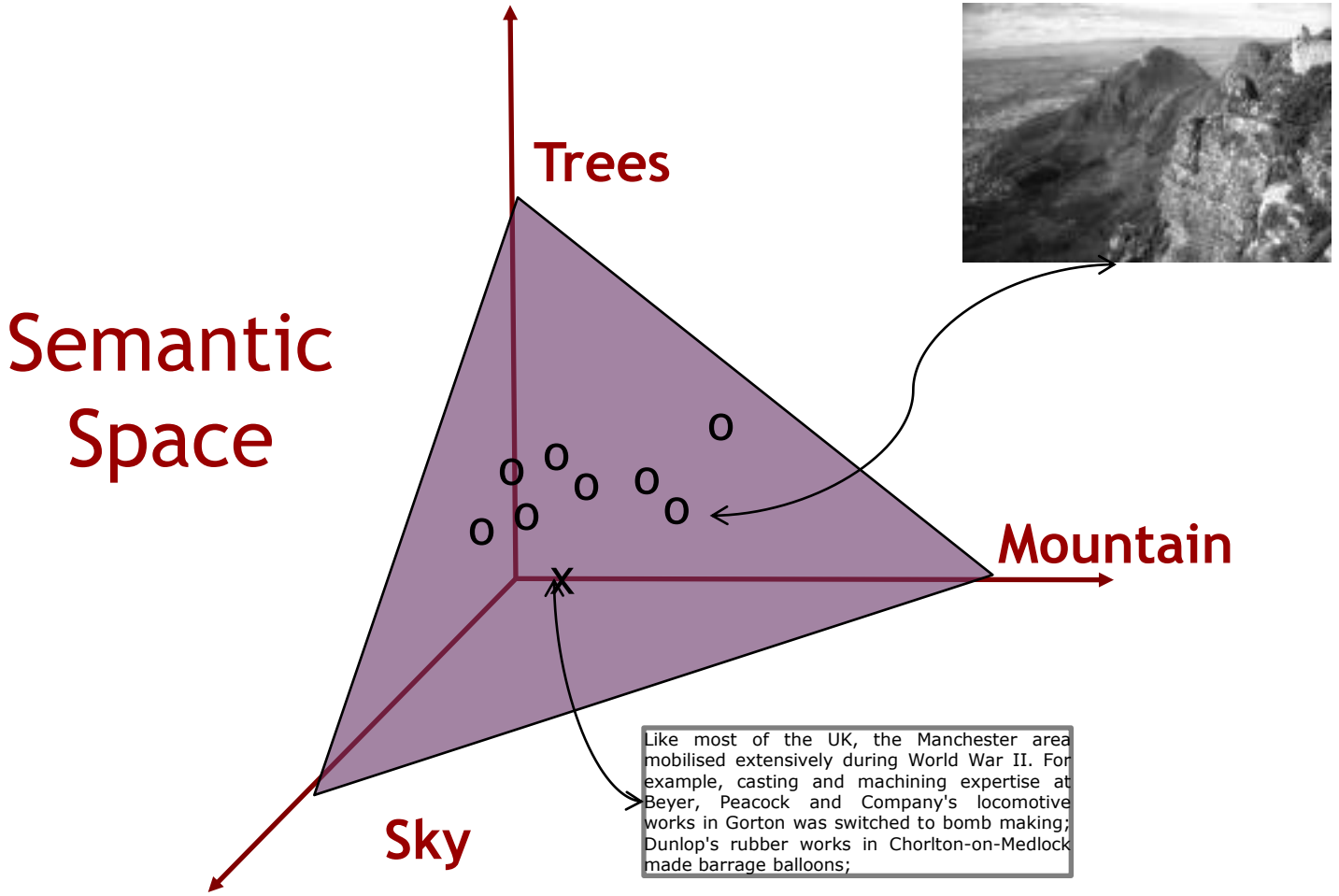
   Can be tested by designing a shared semantic representation for all modalities.

Semantic Matching (SM)

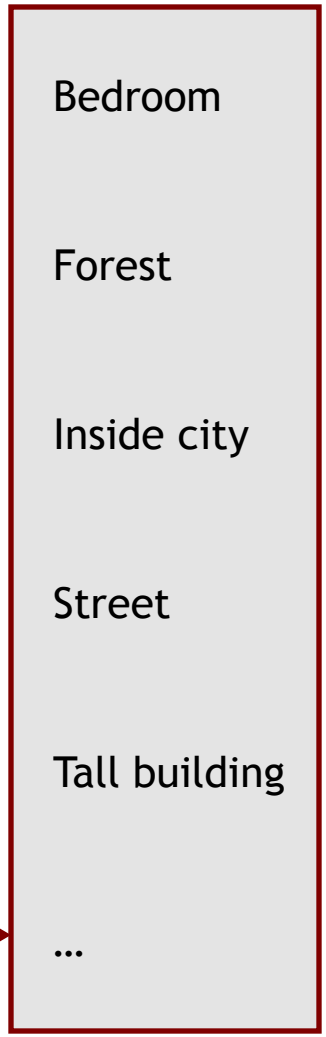
- Design semantic spaces for both modalities [Rasiwasia'07, Smith'03]
  - A space where each dimension is a semantic concept.
  - Each point on this space is a weight vector over these concepts



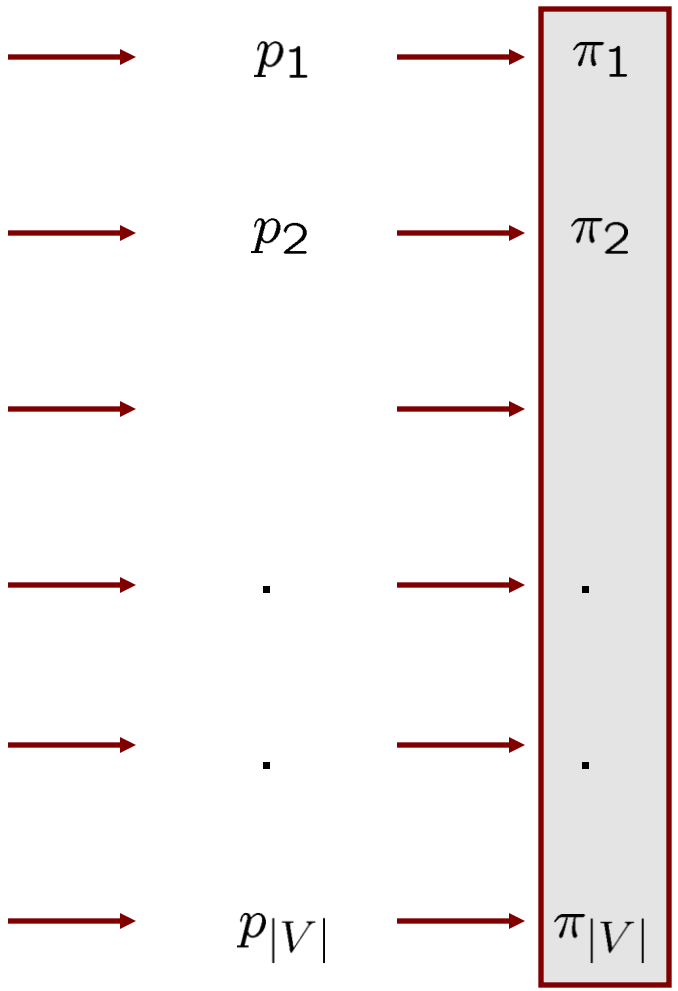
Representation using semantic space

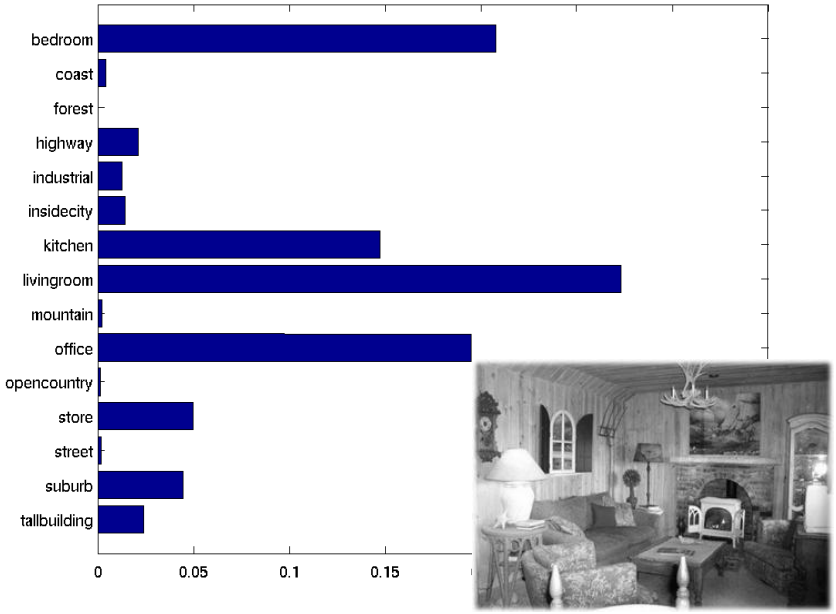


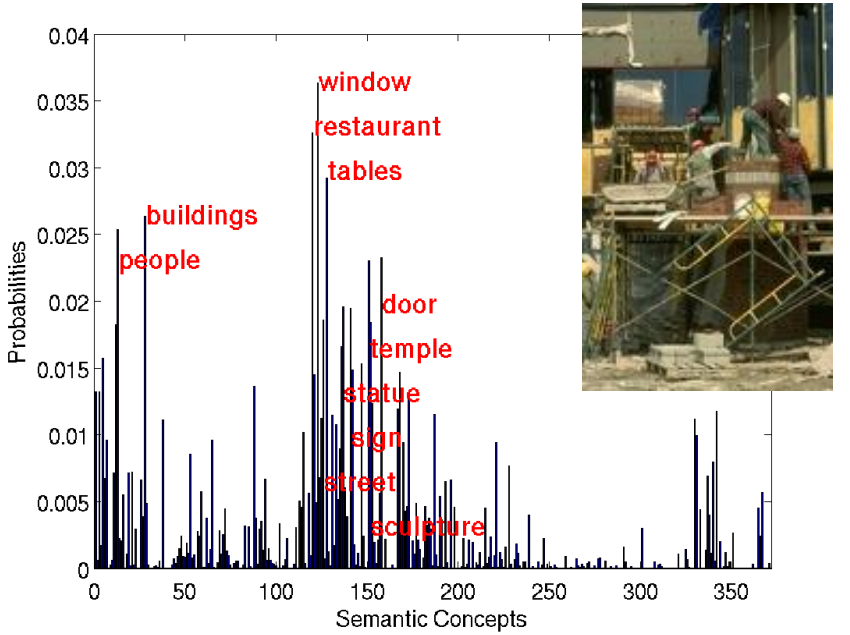
Semantic Representation

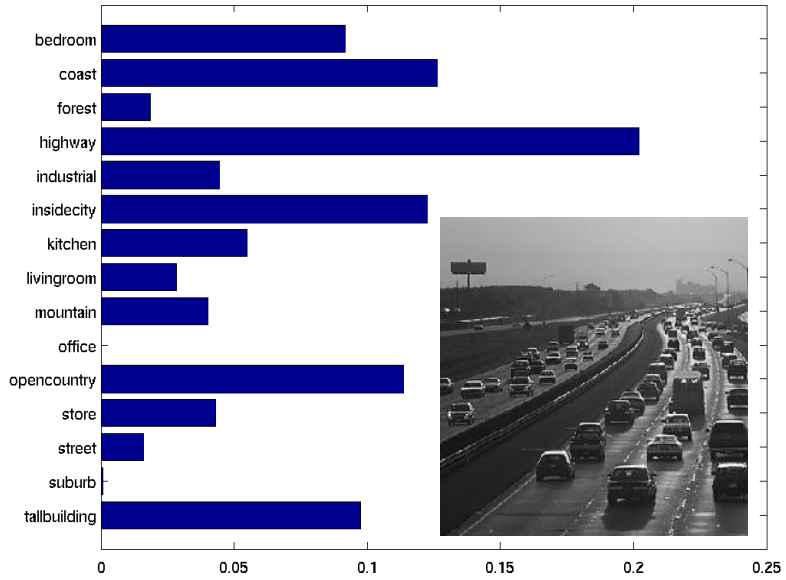


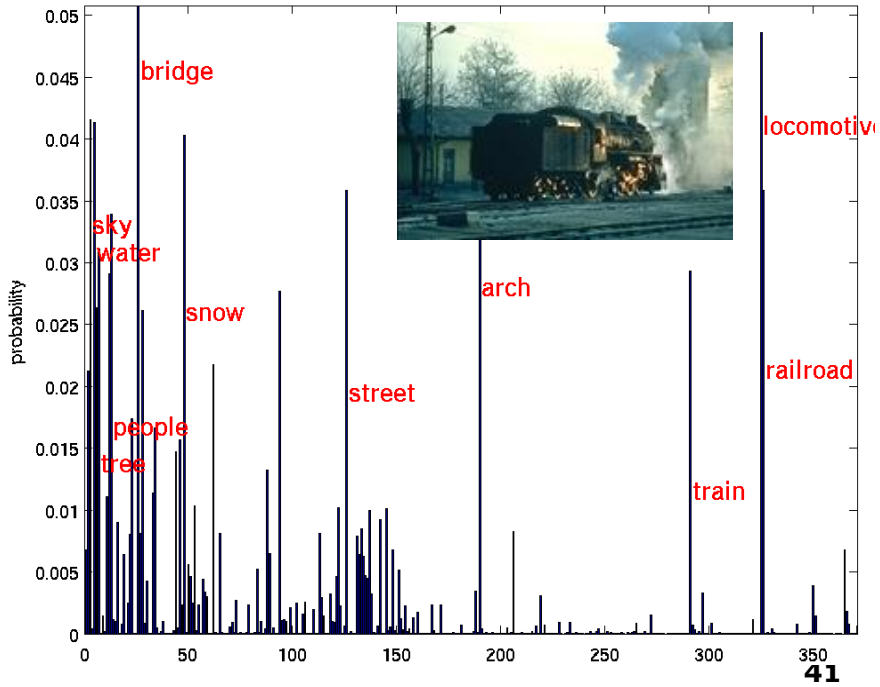
CLASSIFIERS









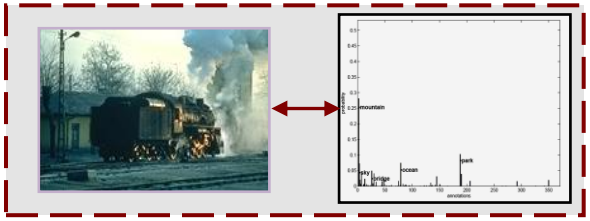


Retrieval on Semantic Space

- Note that Semantic representation is a probability distribution.
- Using suitable similarity function we can compare two vectors.
- For example, one measure is the KL Divergence.

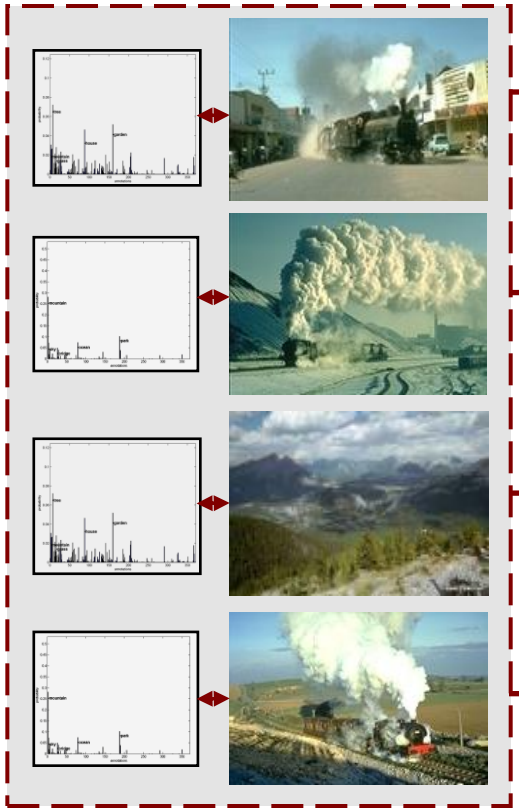
$$f_1(\pi) = \arg\min_y d(\pi, \pi^y)$$

$$d(\pi, \pi') = KL(\pi || \pi') = \sum_{i=1}^{L} \pi_i \log \frac{\pi_i}{\pi'_i}$$



Query

KL divergence



Database

KL divergence

d_1

d_2

d_N

Example retrieval [Rasiwasia '07]

Query

Commercial Construction

People	0.09
Buildings	0.07
Street	0.07
Statue	0.05
Tables	0.04
Water	0.04
Restaurant	0.04

Semantic Features

People	0.08
Statue	0.07
Buildings	0.06
Tables	0.05
Street	0.05
Restaurant	0.04
House	0.03

Buildings	0.06
People	0.06
Street	0.06
Statue	0.04
Tree	0.04
Boats	0.04
Water	0.03

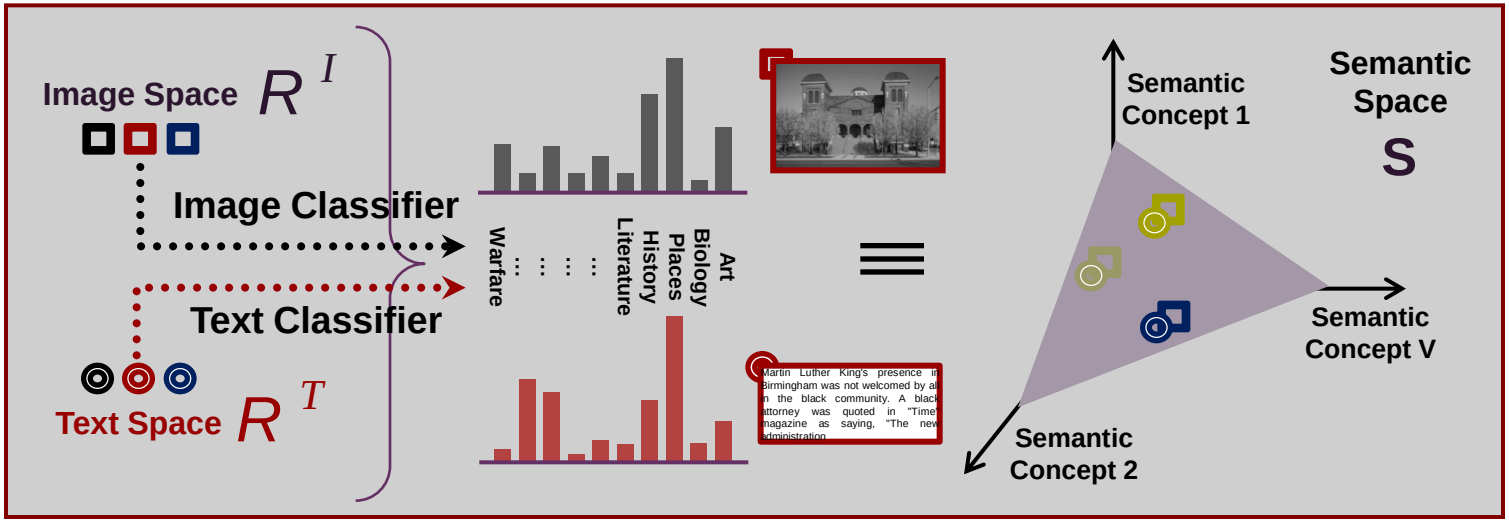
People	0.1
Statue	0.08
Buildings	0.07
Tables	0.06
Street	0.06
Door	0.05
Restaurant	0.04

People	0.12
Restaurant	0.07
Sky	0.06
Tables	0.06
Street	0.05
Buildings	0.05
Statue	0.05

Visual Features

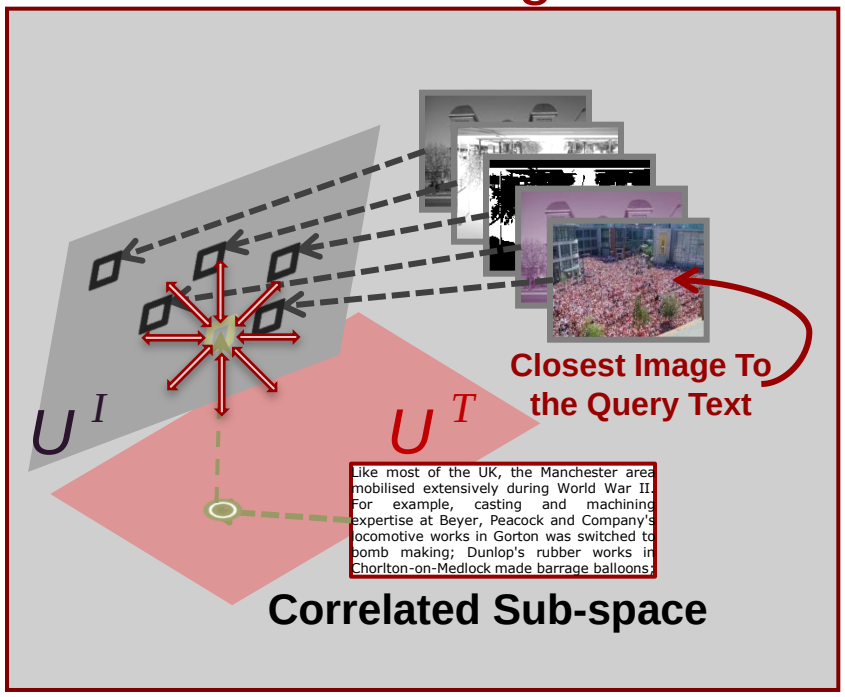
Semantic Matching (SM)

- Design semantic spaces for both modalities [Rasiwasia'07, Smith'03]
  - A space where each dimension is a semantic concept.
  - Each point on this space is a weight vector over these concepts

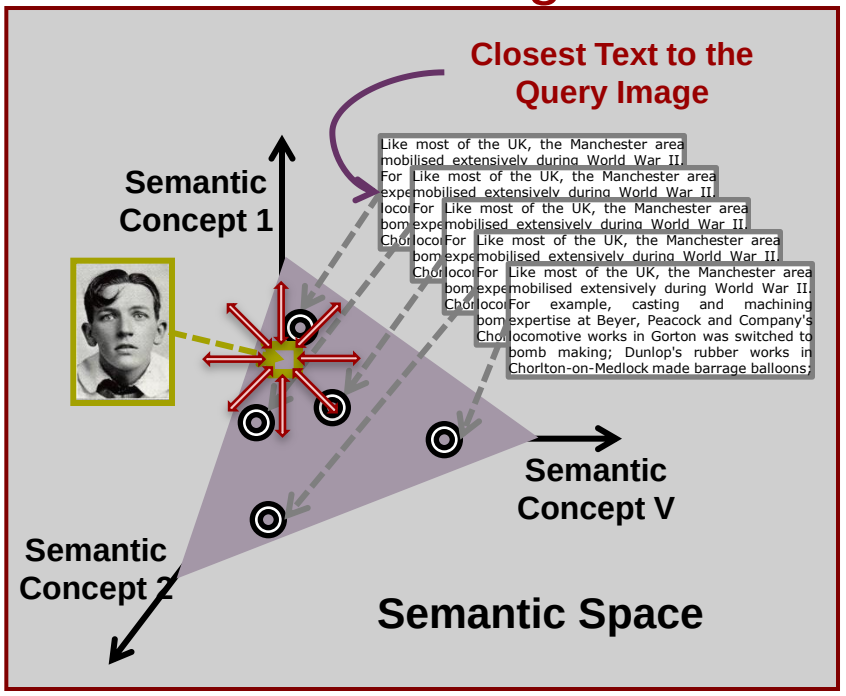


Cross Modal Retrieval

Example: Image to text retrieval using CM



Example: Text to images retrieval using CM



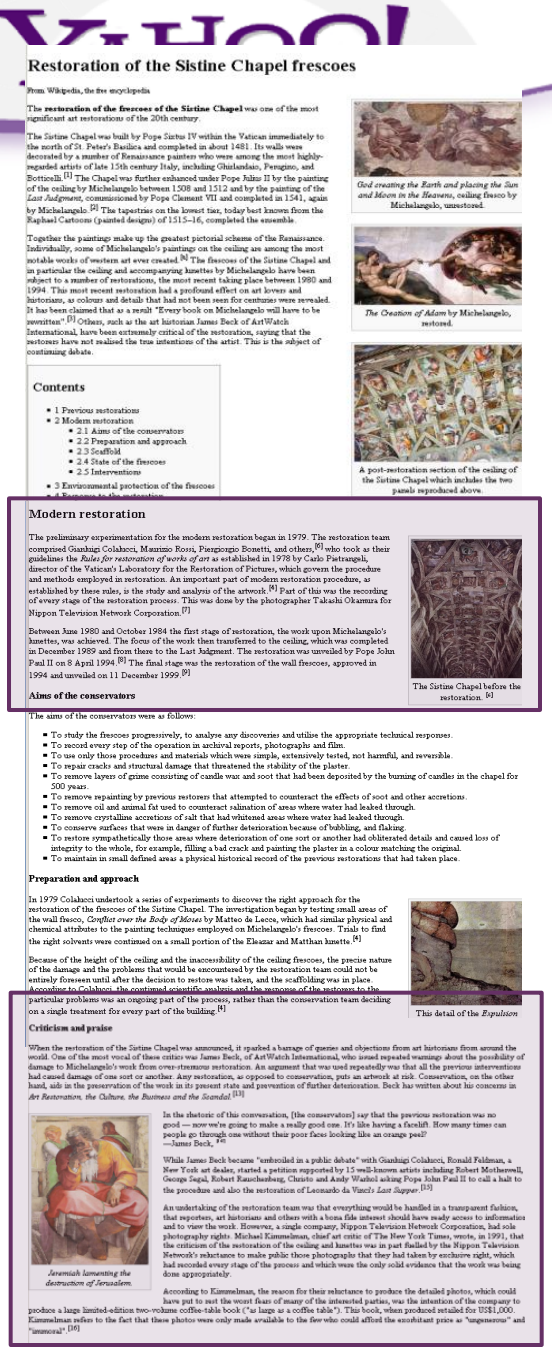
- Ranking is based on a suitable similarity function
  - L2 distance, L1 distance, Normalized Correlation, KL divergence (for SM only) etc.

Experimental Evaluation

- Datasets
- Evaluation Protocol
- Feature Representation
- Parameter Validation
- Results
- Surprise!

Dataset

- Novelty of the problem makes it hard to find existing dataset
- We propose a dataset build using Wikipedia's featured articles
  - 2700 articles, selected and reviewed by Wikipedia's editors since 2009.
  - The articles are accompanied by one or more pictures from the Wikimedia Commons
  - Each article is split into sections that may or may not have an assigned image (sections without images were dropped)
  - Each article is categorized into one of 29 categories (only the 10 most populated categories were chosen)
  - Each 'document' in the proposed set is a 'section of Wikipedia featured article' and its 'associated image'.



Restoration of the Sistine Chapel frescoes

From Wikipedia, the free encyclopedia

The **restoration of the frescoes of the Sistine Chapel** was one of the most significant art restorations of the 20th century.

The Sistine Chapel was built by Pope Sixtus IV within the Vatican immediately to the north of St. Peter's Basilica and completed in about 1481. Its walls were decorated by a number of Renaissance painters who were among the most highly-regarded artists of late 15th century Italy, including Ghirlandaio, Perugino, and Botticelli.[1] The Chapel was further enhanced under Pope Julius II by the painting of the ceiling by Michelangelo between 1508 and 1512 and by the painting of the *Last Judgement*, commissioned by Pope Clement VII and completed in 1541, again by Michelangelo.[2] The tapestries on the lowest tier, today best known from the Raphael Cartoons (painted designs) of 1515–16, completed the ensemble.

Together the paintings make up the greatest pictorial scheme of the Renaissance. Individually, some of Michelangelo's paintings on the ceiling are among the most notable works of western art ever created.[3] The frescoes of the Sistine Chapel and in particular the ceiling and accompanying lunettes by Michelangelo have been subject to a number of restorations, the most recent taking place between 1980 and 1994. This most recent restoration had a profound effect on art lovers and historians, as colours and details that had not been seen for centuries were revealed. It has been claimed that as a result "Every book on Michelangelo will have to be rewritten".[3] Others, such as the art historian James Beck of ArtWatch International, have been extremely critical of the restoration, saying that the restorers have not realised the true intentions of the artist. This is the subject of continuing debate.

God creating the Earth and placing the Sun and Moon in the Heavens, ceiling fresco by Michelangelo, unrestored.

The Creation of Adam by Michelangelo, restored.

A post-restoration section of the ceiling of the Sistine Chapel which includes the two panels reproduced above

Contents

- 1 Previous restorations
- 2 Modern restoration
  - 2.1 Aims of the conservators
  - 2.2 Preparation and approach
  - 2.3 Scaffold
  - 2.4 State of the frescoes
  - 2.5 Interventions
- 3 Environmental protection of the frescoes
- 4 Economy to the restoration

Modern restoration

The preliminary experimentation for the modern restoration began in 1979. The restoration team comprised Gianluigi Colalucci, Maurizio Rossi, Piergiorgio Bonetti, and others,[6] who took as their guideline the *Rules for restoration of works of art* as established in 1978 by Carlo Pietrangeli, director of the Vatican's Laboratory for the Restoration of Pictures, which govern the procedure and methods employed in restoration. An important part of modern restoration procedure, as established by these rules, is the study and analysis of the artwork.[6] Part of this was the recording of every stage of the restoration process. This was done by the photographer Takashi Okamura for Nippon Television Network Corporation.[7]

Between June 1980 and October 1984 the first stage of restoration, the work upon Michelangelo's lunettes, was achieved. The focus of the work then transferred to the ceiling, which was completed in December 1989 and from there to the Last Judgment. The restoration was unveiled by Pope John Paul II on 8 April 1994.[8] The final stage was the restoration of the wall frescoes, approved in 1994 and unveiled on 11 December 1999.[9]

The Sistine Chapel before the restoration.[6]

Aims of the conservators

The aims of the conservators were as follows:

- To study the frescoes progressively, to analyse any discoveries and utilize the appropriate technical responses.
- To record every step of the operation in archival reports, photographs and film.
- To use only those procedures and materials which were simple, extensively tested, not harmful, and reversible.
- To repair cracks and structural damage that threatened the stability of the plaster.
- To remove layers of grime consisting of candle wax and soot that had been deposited by the burning of candles in the chapel for 500 years.
- To remove repainting by previous restorers that attempted to counteract the effects of soot and other accretions.
- To remove oil and animal fat used to counteract salination of areas where water had leaked through.
- To remove crystalline accretions of salt that had whitened areas where water had leaked through.
- To conserve surfaces that were in danger of further deterioration because of bubbling, and flaking.
- To restore sympathetically those areas where deterioration of one sort or another had obliterated details and caused loss of integrity to the whole, for example, filling a bad crack and painting the plaster in a colour matching the original.
- To maintain in small defined areas a physical historical record of the previous restorations that had taken place.

Preparation and approach

In 1979 Colalucci undertook a series of experiments to discover the right approach for the restoration of the frescoes of the Sistine Chapel. The investigation began by testing small areas of the wall fresco, *Conflict over the Body of Moses* by Matteo de Lecce, which had similar physical and chemical attributes to the painting techniques employed on Michelangelo's frescoes. Trials to find the right solvents were continued on a small portion of the Eleazar and Matthan lunette.[8]

This detail of the *Expulsion*

Because of the height of the ceiling and the inaccessibility of the ceiling frescoes, the precise nature of the damage and the problems that would be encountered by the restoration team could not be entirely foreseen until after the decision to restore was taken, and the scaffolding was in place. According to Colalucci, the continued scientific analysis and the progress of the restorers to the particular problems was an ongoing part of the process, rather than the conservation team deciding on a single treatment for every part of the building.[6]

Criticism and praise

When the restoration of the Sistine Chapel was announced, it sparked a barrage of queries and objections from art historians from around the world. One of the most vocal of these critics was James Beck, of ArtWatch International, who issued repeated warnings about the possibility of damage to Michelangelo's work from over-strenuous restoration. An argument that was used repeatedly was that all the previous interventions had caused damage of one sort or another. Any restoration, as opposed to conservation, puts an artwork at risk. Conservation, on the other hand, aids in the preservation of the work in its present state and prevention of further deterioration. Beck has written about his concerns in *Art Restoration, the Culture, the Business and the Scandal*.[13]

Jeremiah lamenting the destruction of Jerusalem.

In the rhetoric of this conversation, [the conservators] say that the previous restoration was no good — now we're going to make a really good one. It's like having a facelift. How many times can people go through one without their poor faces looking like an orange peel?
—James Beck, [14]

While James Beck became "embroiled in a public debate" with Gianluigi Colalucci, Ronald Feldman, a New York art dealer, started a petition supported by 15 well-known artists including Robert Motherwell, George Segal, Robert Rauschenberg, Christo and Andy Warhol asking Pope John Paul II to call a halt to the procedure and also the restoration of Leonardo da Vinci's *Last Supper*.[15]

An undertaking of the restoration team was that everything would be handled in a transparent fashion, that reporters, art historians and others with a bona fide interest should have ready access to information and to view the work. However, a single company, Nippon Television Network Corporation, had sole photography rights. Michael Kimmelman, chief art critic of The New York Times, wrote, in 1991, that the criticism of the restoration of the ceiling and lunettes was in part fuelled by the Nippon Television Network's reluctance to make public those photographs that they had taken by exclusive right, which had recorded every stage of the process and which were the only solid evidence that the work was being done appropriately.

According to Kimmelman, the reason for their reluctance to produce the detailed photos, which could have put to rest the worst fears of many of the interested parties, was the intention of the company to produce a large limited-edition two-volume coffee-table book ("as large as a coffee table"). This book, when produced retailed for US$1,000. Kimmelman refers to the fact that these photos were only made available to the few who could afford the exorbitant price as "ungenerous" and "immoral".[16]

Dataset characterization

Category	Training	Query/ Retrieval	Total documents
Art & architecture	138	34	172
Biology	272	88	360
Geography & places	244	96	340
History	248	85	333
Literature & theatre	202	65	267
Media	178	58	236
Music	186	51	237
Royalty & nobility	144	41	185
Sport & recreation	214	71	285
Warfare	347	104	451
total	2173	693	2866

Geography and Places

The population of Turkey stood at 71.5 million with a growth rate of 1.31% per annum, based on the 2008 Census. It has an average population density of 92 persons per km². The proportion of the population residing in urban areas is 70.5%. People within the 15–64 age group constitute 66.5% of the total population, the 0–14 age group corresponds 26.4% of the population, while 65 years and higher of age correspond to 7.1% of the total population. Life expectancy stands at 70.67 years for men and 75.73 years for women, with an overall average of 73.14 years for the populace as a whole. Education is compulsory and free from ages 6 to 15. The literacy rate is 95.3% for men and 79.6% for women, with an overall average of 87.4%. The low figures for women are mainly due to the traditional customs of the Arabs and Kurds who live in the southeastern provinces of the country. Article 66 of the Turkish Constitution defines a "Turk" as "anyone who is bound to the Turkish state through the bond of citizenship"; therefore, the legal use of the term "Turkish" as a citizen of Turkey is different from the ethnic definition. (…)

A number of variants were built on the same chassis as the TAM tank. The original program called for the design of an infantry fighting vehicle, and in 1977 the program finished manufacturing the prototype of the ''Vehículo de Combate Transporte de Personal'' (Personnel Transport Combat Vehicle), or VCTP. The VCTP is able to transport a squad of 12 men, including the squad leader and nine riflemen. The squad leader is situated in the turret of the vehicle; one rifleman sits behind him and another six are seated in the chassis, the eighth manning the hull machine gun and the ninth situated in the turret with the gunner. All personnel can fire their weapons from inside the vehicle, and the VCTP's turret is armed with Rheinmetall's Rh-202 20 millimeter (.79 in) autocannon. The VCTP holds 880 rounds for the autocannon, including subcaliber armor-piercing DM63 rounds. It is also armed with a 7.62 millimeter FN MAG 60-20 mounted on the turret roof. Infantry can dismount through a door on the rear of the hull. (…)

Warfare

Culture and Society

Despite agreeing on most issues regarding the protection of national parks, friction between the NPA and NPS was seemingly unavoidable. Mather and Yard disagreed on many issues; whereas Mather was not interested in the protection of wildlife and accepted the Biological Survey's efforts to exterminate predators within parks, Yard vehemently criticized the program as early as 1924 (Fox, p. 204). Yard was also highly critical of Mather's administration of the parks. Mather advocated plush accommodations, city comforts and various entertainments to encourage park visitation. These plans clashed with Yard's ideals, and he considered such urbanization of the nation's parks misguided. While visiting Yosemite National Park in 1926, he stated that the valley was "lost" after finding crowds, automobiles, jazz music and even a bear show (Sutter, p. 126). In 1924, the United States Forest Service initiated a program to set aside "primitive areas" in the national forests that protected wilderness while opening it to use. (…)

TVGraz Dataset

- A collection of webpages compiled by [Khan etal, Workshop of the Austrian Association for Pattern Recognition'09]
- Used images search to obtain the top 1000 results for ten category names from the Caltech-256 dataset.
- The results were filtered into a set of 2592 webpages
- Because of copyright issues only a list of URLs is provided
- Our compiled version has only 2058 pairs of image and text, due to our removal of defunct URLs.

http://blog.thomaslaupstad.com/2007/07/23/close-up-photo-of-a-common-frog/

Close up photo of a common frog

July 23rd, 2007 · 6 Comments

Taken with Olympus E500 digital camera on July 15th, 2007. Click picture for larger view.

This is a close up photo of a common frog (*Rana temporaria*) being very camera shy. I found the frog pretty far up in the mountains and I was surprised because I have never seen frogs in this area before.

Frogs are supposed to be an biological indicator that are able to tell us whether the environment is healthy or unhealthy. The reason for this is that frogs are very sensitive to pollution and habitat degradation.
I'm very happy that I found this frog because it means the nature around here is healthy

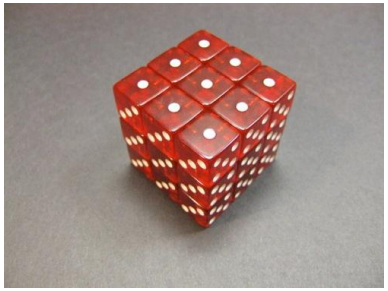
Dataset characterization

Category	Training	Query/ Retrieval	Total documents
Brain	109	47	158
Butterfly	195	51	246
Cactus	137	37	174
Deer	223	51	274
Dice	169	50	219
Dolphin	163	59	222
Elephant	120	54	174
Frog	215	67	282
Harp	131	42	173
Pram	96	42	138
total	1558	500	2058

Frog

This is a close up photo of a common frog (*Rana temporaria*) being very camera shy. I found the frog pretty far up in the mountains and I was surprised because I have never seen frogs in this area before.
Frogs are supposed to be an biological indicator that are able to tell us whether the environment is healthy or unhealthy. The reason for this is that frogs are very sensitive to pollution and habitat degradation.
I'm very happy that I found this frog because it means the nature around here is healthy
If you view the picture large you can actually see an ant below the frog on left side. Maybe it was the frogs next meal?

Burzvingion made a really nice dice based magnetic Rubik's dice cube
"The non-symetrical nature of some of the die patterns gives the dice cube an extra element of strategy not present in regular Rubik's Cubes. This variable of rotation is an aspect of the "sudocube," but the dice cube is easier to solve, because each face contains identical figures, and every cubie is unique--two properties of the original Rubik's Cube that are not carried over to the "sudocube.""



Dice

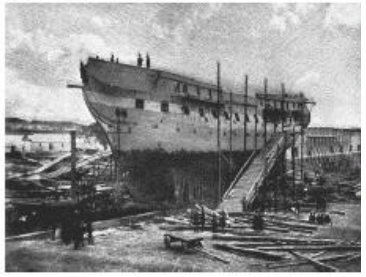
Elephant

When visiting Chiang Mai in the north of Thailand you can't seem to walk more than 100 metres without passing a shop selling elephant riding tours.
There are many elephant riding trips available packaged with other activities, usually hill tribe treks and some other activity like white water rafting. These sort of day tours make for good value when you consider how much time you would save if you went to each activity individually.
Most elephant rides go for half an hour to one hour. There is no need to go for over an hour, unless you really love elephants.

Text to Image Query (1)

Between October 1 and October 17, the Japanese delivered 15,000 troops to Guadalcanal, giving Hyakutake 20,000 total troops to employ for his planned offensive. Because of the loss of their positions on the east side of the Matanikau, the Japanese decided that an attack on the U.S. defenses along the coast would be prohibitively difficult. Therefore, Hyakutake decided that the main thrust of his planned attack would be from south of Henderson Field. His 2nd Division (augmented by troops from the 38th Division), under Lieutenant General Masao Maruyama and comprising 7,000 soldiers in three infantry regiments of three battalions each was ordered to march through the jungle and attack the American defences from the south near the east bank of the Lunga River. The date of the attack was set for October 22, then changed to October 23. To distract the Americans from the planned attack from the south, Hyakutake's heavy artillery plus five battalions of infantry (about 2,900 men) under Major General Tadashi Sumiyoshi were to attack the American defenses from the west along the coastal corridor. The Japanese estimated that there were 10,000 American troops on the island, when in fact there were about 23,000…

Top 5 Retrieved Images


Text to Image Query (2)

Around 850, out of obscurity rose Vijayalaya, made use of an opportunity arising out of a conflict between Pandyas and Pallavas, captured Thanjavur and eventually established the imperial line of the medieval Cholas. Vijayalaya revived the Chola dynasty and his son Aditya I helped establish their independence. He invaded Pallava kingdom in 903 and killed the Pallava king Aparajita in battle, ending the Pallava reign. K.A.N. Sastri, ''A History of South India'' p 159 The Chola kingdom under Parantaka I expanded to cover the entire Pandya country. However towards the end of his reign he suffered several reverses by the Rashtrakutas who had extended their territories well into the Chola kingdom…

Top 5 Retrieved Images





Text to Image Query (3)

The lumber boom on Plunketts Creek ended when the virgin timber ran out. By 1898, the old growth hemlock was exhausted and the Proctor tannery, then owned by the Elk Tanning Company, was closed and dismantled. Lumbering continued in the watershed, but the last logs were floated down Plunketts Creek to the Loyalsock in 1905. The Susquehanna and Eagles Mere Railroad was abandoned in sections between 1922 and 1930, as the lumber it was built to transport was depleted. The CPL logging railroad and their Masten sawmills were abandoned in 1930. Without timber, the populations of Proctor and Barbours declined. The Barbours post office closed in the 1930s and the Proctor post office closed on July 1, 1953. Both villages also lost their schools and almost all of their businesses. Proctor celebrated its centennial in 1968, and a 1970 newspaper article on its thirty-ninth annual "Proctor Homecoming" reunion called it a "near-deserted old tannery town". In the 1980s, the last store in Barbours closed, and the former hotel (which had become a hunting club) was torn down to make way for a new bridge across Loyalsock Creek…

Top 5 Retrieved Images





Text to Image Query (4)

On the Nature Trail behind the Bathabara Church ,there are numerous wild flowers and plants blooming, that attract a variety of insects,bees and birds. Here a beautiful Butterfly is attracted to the blooms of the Joe Pye Weed.

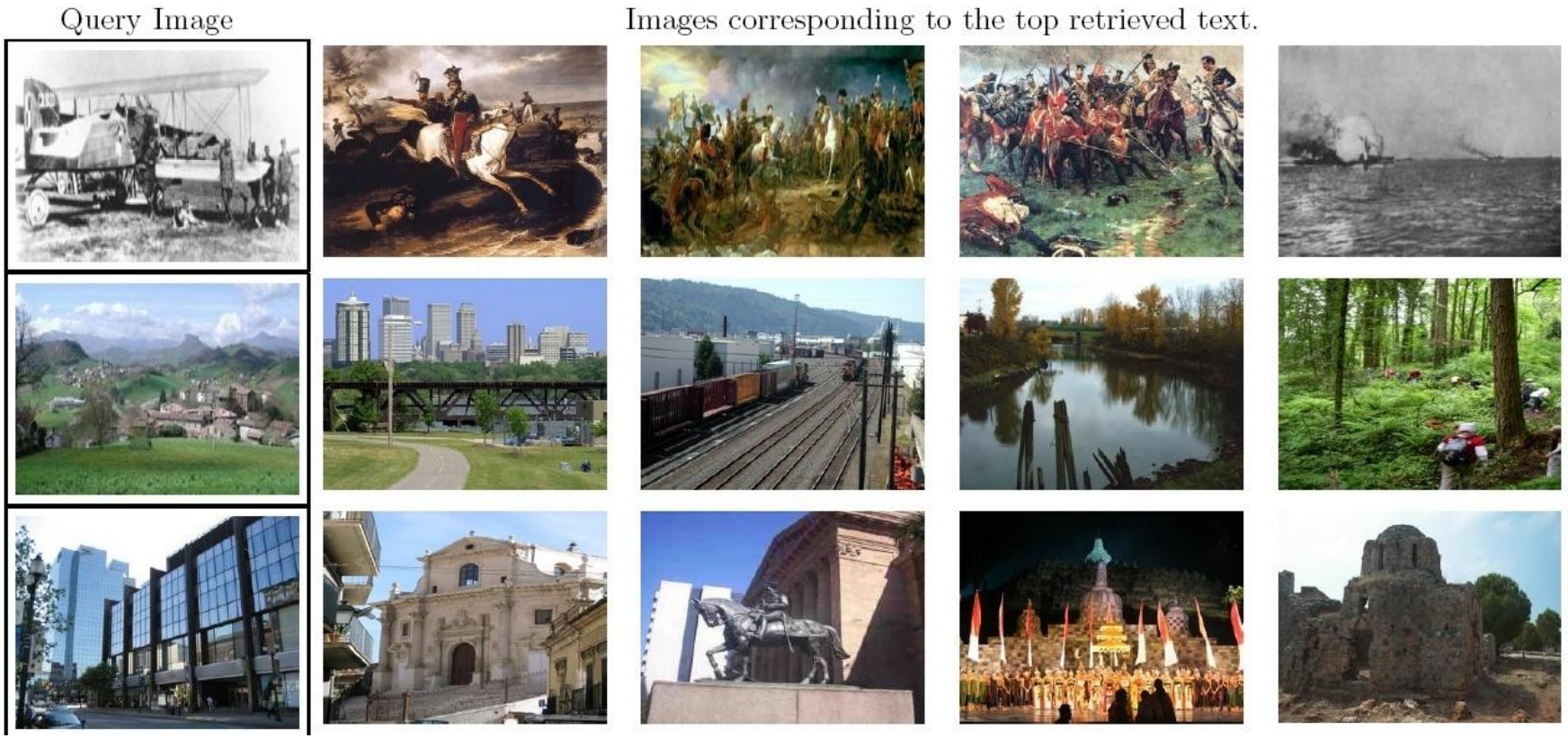
Top 5 Retrieved Images





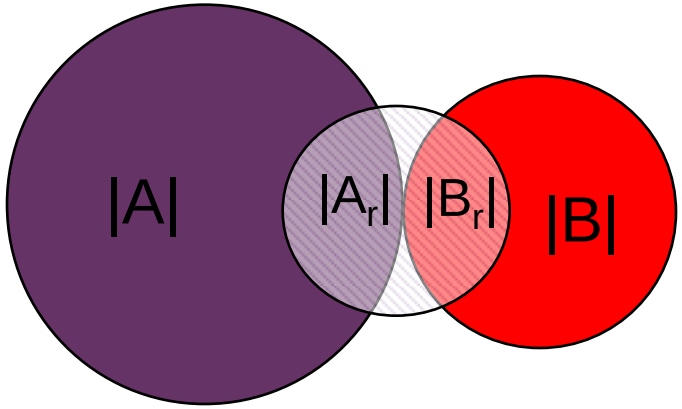
Text to Image Retrieval Example

- Ground truth image corresponding to the retrieved text is shown



Evaluation – Precision:Recall

Precision - Recall



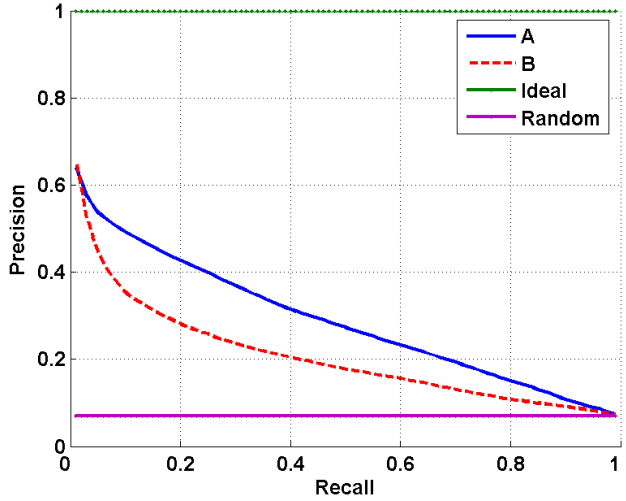
Irrelevant Images |A|

Relevant Images |B|

Retrieved Images $|A_r|+|B_r|$

Precision - Recall

Curves



Precision = $\frac{|B_r|}{|A_r|+|B_r|}$

The proportion of retrieved and relevant images to all the images retrieved.

Recall = $\frac{|B_r|}{|B|}$

The proportion of relevant images that are retrieved, out of all relevant images available.

Retrieval Performance (TVGraz)

- The performance of both Correlation & Semantic Matching is over 400% better than chance.

Mean Average Precision

Model	Image query	Text query	Avg.
Chance	0.114	0.114	0.114
CM	0.507	0.486	**0.497**
SM	0.625	0.618	**0.622**

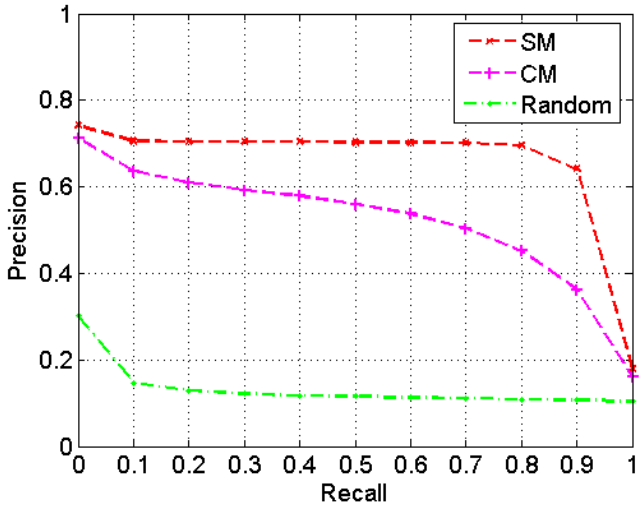
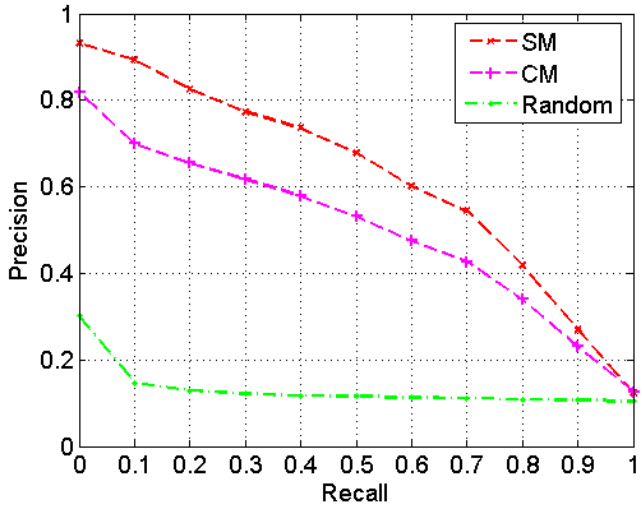


Image to Text



Text to Image

Retrieval Performance (Wikipedia)

- The performance of both Correlation & Semantic Matching is over 100% better than chance.

Mean Average Precision

Model	Image query	Text query	Avg.
Chance	0.119	0.119	0.119
CM	0.282	0.225	**0.253**
SM	0.362	0.252	**0.307**

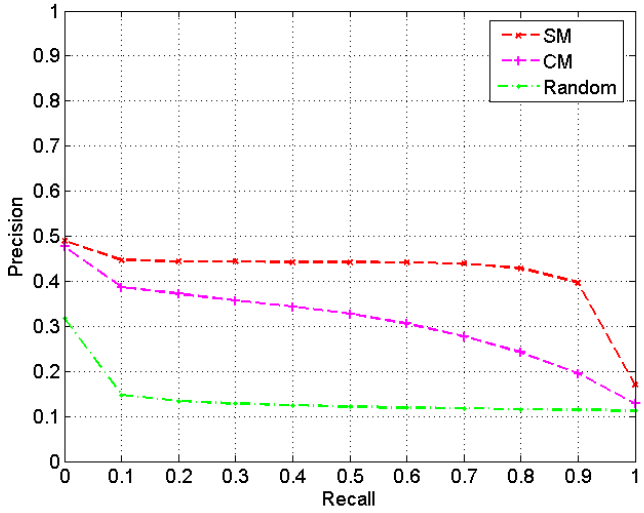
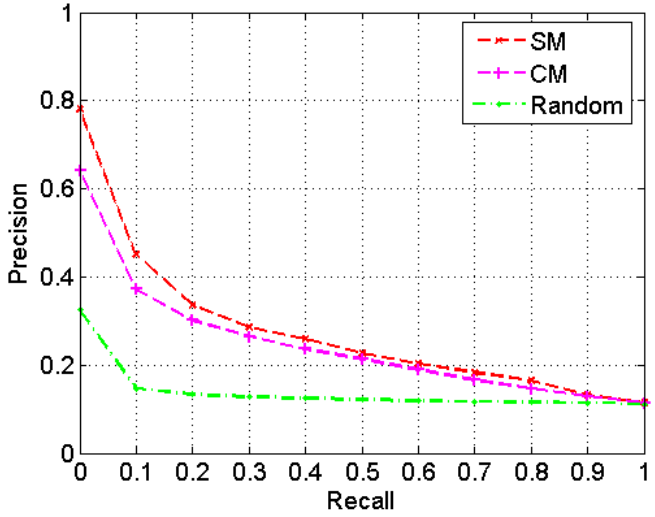


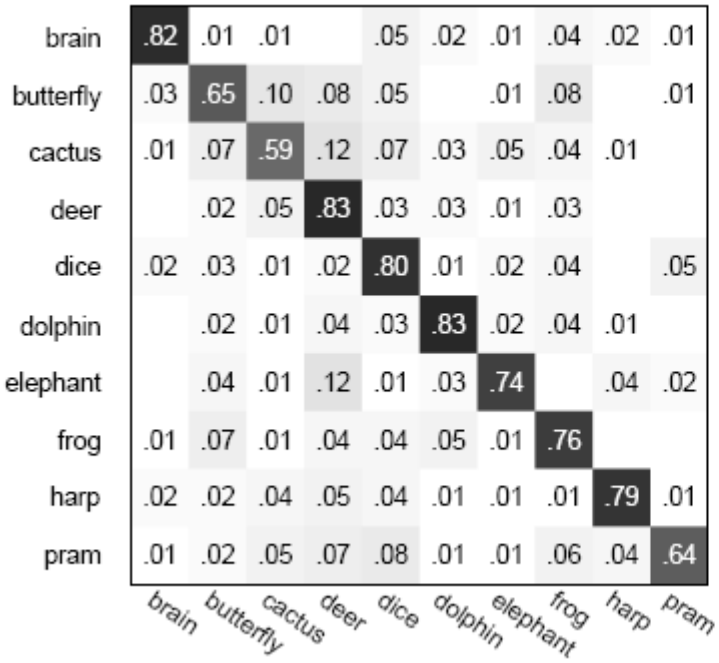
Image to Text



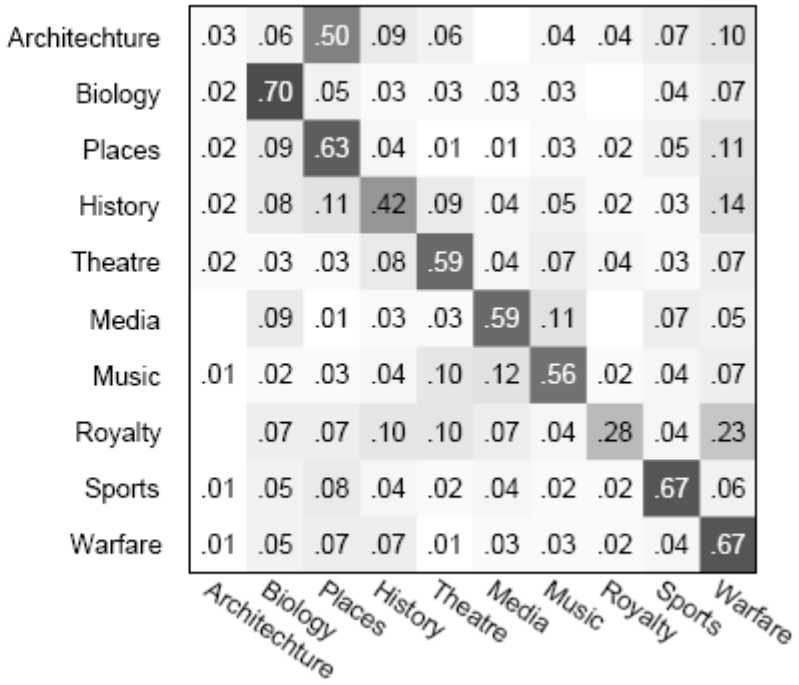
Text to Image

Why a difference?

- Categories for TVGraz are specific and well-seperated, with little overlap.



TVGraz

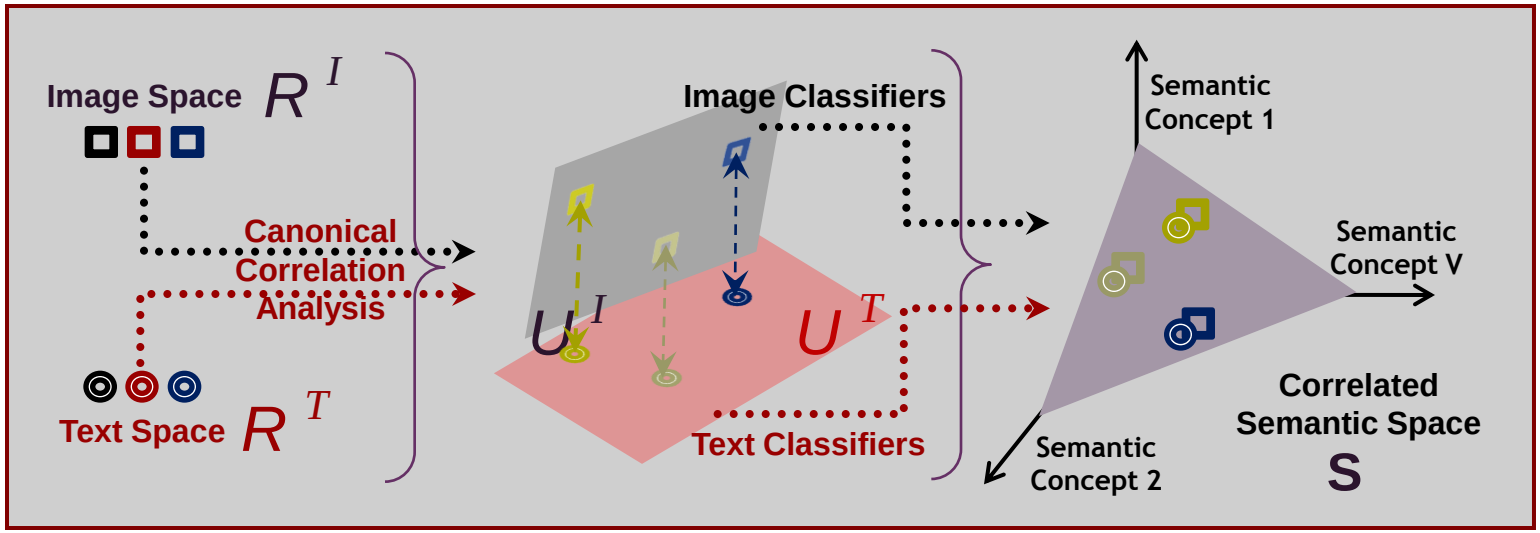


Wikipedia

- Eg: On Wikipedia, Architecture gets confused with Places, History gets confused with Warfare.

Semantic Correlation Matching (SCM)

- *Although CM and SM work on different principles they are not mutually exclusive.*
- Combination of the two approaches can lead to improved performance
  - Learn the maximally-correlated subspaces using CCA
  - Design semantic spaces using the correlated feature as the low-level representation.



Retrieval Performance

- Combining the benefits of CM and SM leads to further improvements on both datasets.

Mean Average Precision

TVGraz

Model	Image query	Text query	Avg.
Chance	0.114	0.114	0.114
CM	0.507	0.486	0.497
SM	0.625	0.618	0.622
SCM	0.697	0.692	**0.694**

Wikipedia

Model	Image query	Text query	Avg.
Chance	0.119	0.119	0.119
CM	0.282	0.225	0.253
SM	0.362	0.252	0.307
SCM	0.377	0.274	**0.325**

Conclusion

- Proposed an approach to build cross-modal retrieval systems.
- Explored two hypotheses
  - CM: The problem is that there is no correlation between the representations of different modalities.
  - SM: The problem is that the representation lacks common semantics.
- Both CM and SM hypotheses holds true
  - Tested by building intermediate spaces based on maximizing correlation and a common semantic representation.
- CM and SM are not mutually exclusive and their combination leads to further improvements.

Questions?

Leveraging Text for Image Retrieval

- Suppose the task is uni-modal – to retrieve images using an image query.

- Furthermore, if the images in the retrieval set are supported with text.

- Then, is there any benefit in cross-modal retrieval?
  - Effectively ignoring image data.

Leveraging Text for Image Retrieval

- Preliminary results are positive.

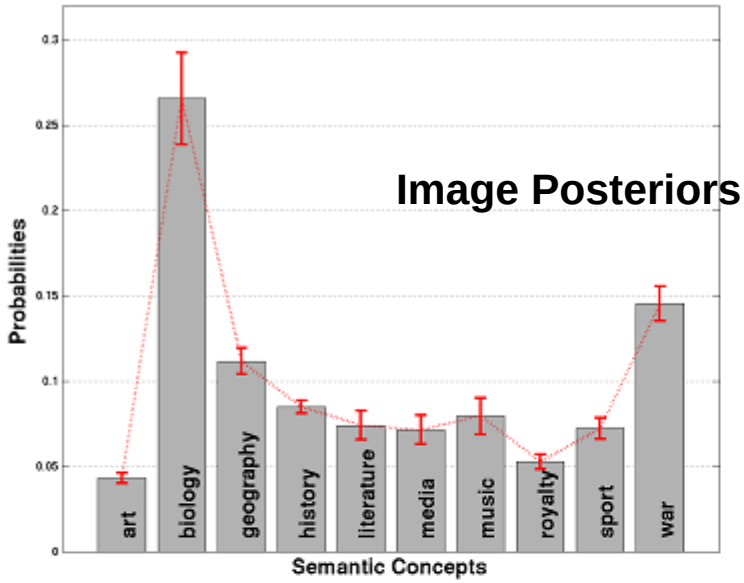
Mean Average Precision

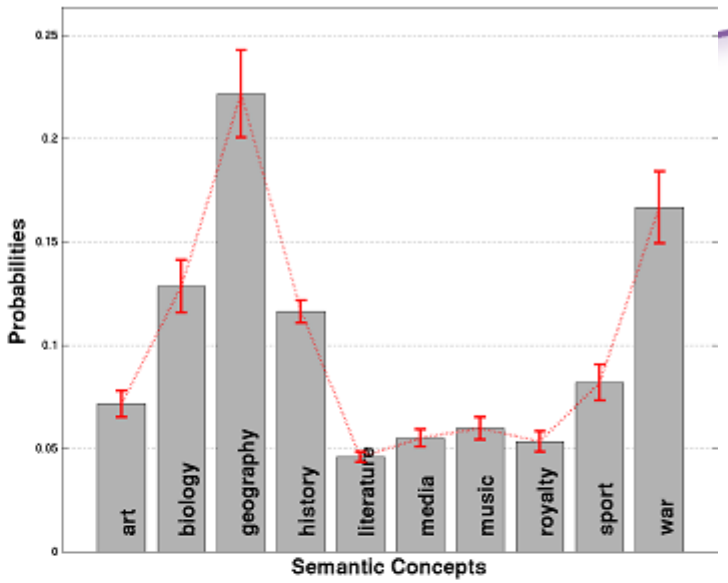
Model	TVGraz	Wikipedia
Cross-Modal	**0.624**	**0.358**
Uni-modal (Semantic)	0.432	0.162
Uni-modal (Histogram)	0.238	0.139
Random	0.111	0.117

Some examples

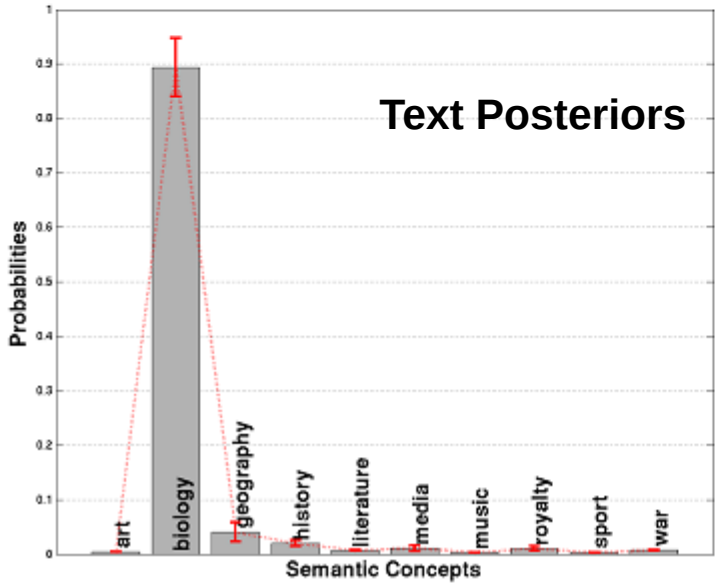
Top rows shows retrieved results using Uni-modal

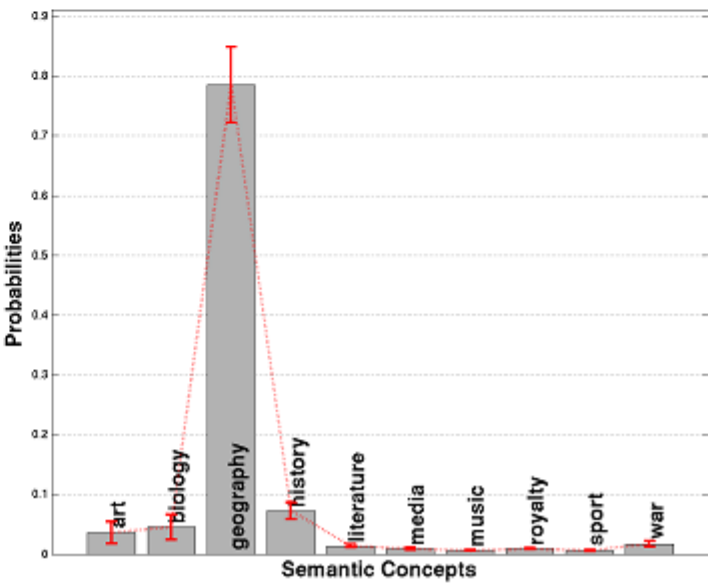
Image Posteriors





Text Posteriors



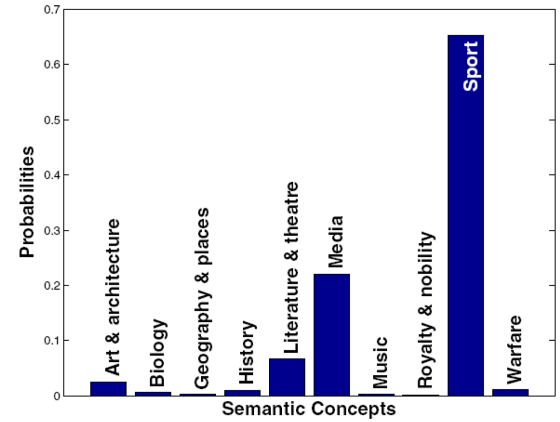


Intuitions

- Intrinsic ambiguity for images is higher.
- There exists other modalities that have less ambiguity, such as Text
- Cross modal search can leverage on this to produce improved uni-modal search results.

…future work.

Frank J. Selke served as Chairman of the selection committee from 1960 until 1971, when he resigned because of the induction of Harvey "Busher" Jackson. Jackson, known for his o↑-ice lifestyle, had died in 1966 of liver failure. Selke would not condone the induction and even tried to block it because he considered Jackson a poor role model."Honoured members: the Hockey Hall of Fame", p. 91
On March 30, 1993, it was announced that Gil Stein, who at the time was the president of the National Hockey League, would be inducted into the Hall of Fame. There were immediate allegations that he had engineered his election through manipulation of the hall's board of directors. Due to these allegations, NHL commissioner Gary Bettman hired two independent lawyers, Arnold Burns and Yves Fortier, to lead an investigation. They concluded that Stein had "improperly manipulated the process" and "created the false appearance and illusion" that his nomination was the idea of Bruce McNall. They concluded that Stein pressured McNall to nominate him and had refused to withdraw his nomination when asked to do so by Bettman. There was a dispute over McNall's role and Stein was "categorical in stating that the idea was Mr. McNall's." They recommended that Stein's selection be overturned, but it was revealed Stein had decided to turn down the induction before their announcement.
In 1989, Alan Eagleson, a long time executive director of the National Hockey League builder. He resigned nine years later from the Hall after pleading guilty to mail fraud of dollars from the NHL Players Association pension funds."Honoured members: tl resignation came six days before a vote was scheduled to determine if he should be e Hall of Fame was not going to become involved in the issue, but was forced to act whe Orr, Ted Lindsay and Brad Park, campaigned for Eagleson's expulsion, even threate he was not removed. He became the ↓rst member of a sports hall of fame in North A

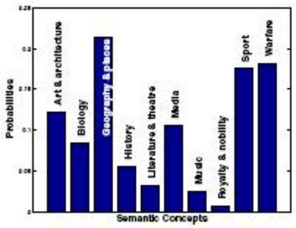


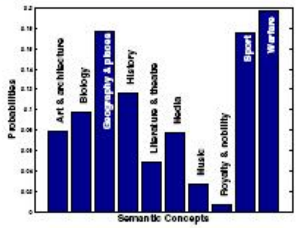
Text to image Retrieval

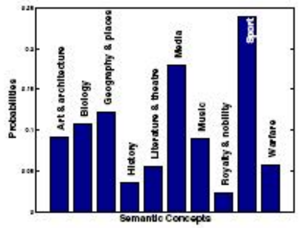
Query Text and its semantic vector

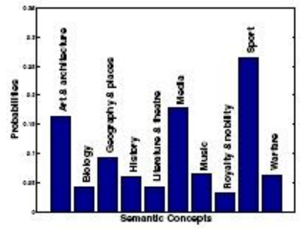
Retrieved Images and their semantic vectors

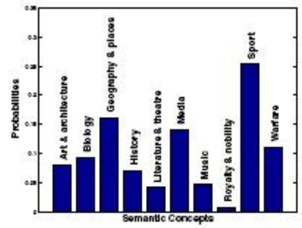


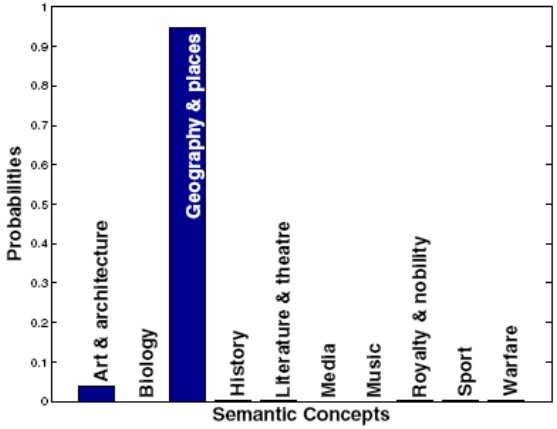








The pre-collegiate medium of instruction in schools is predominantly Kannada, while English and Kannada are predominant languages in private schools. Additionally, other media of instruction exist in Mangalore. The medium of instruction in educational institutions after matriculation in colleges is English. Recently, a committee of experts constituted by the Tulu Sahitya Academy recommended the inclusion of Tulu (in Kannada script) as a medium of instruction in education.
Schools and colleges in Mangalore are either government-run or run by private trusts and individuals. The schools are a¡liated with either the Karnataka State Board, Indian Certi⌊cate of Secondary Education (ICSE), or the Central Board for Secondary Education (CBSE) boards. After completing 10 years of schooling in secondary education, students enroll in Higher Secondary School, specializing in one of the three streams â Arts, Commerce or Science. Since the 1980s, there have been a large number of professional institutions established in a variety of ⌊elds including engineering, medicine, dentistry, business management and hotel management. The earliest schools established in Mangalore were the Basel Evangelical School (1838) and Milagres School (1848). The Kasturba Medical College establ medical college. Popular educational institutions in the city are National Institute Medical Academy,Father Muller Medical College, St. Aloysius College, Canara Coll College and St. Joseph Engineering College. The Bibliophile's Paradise, a hi-tech Bank, is located at Mannagudda in Mangalore. Mangalore University was established higher educational needs of Dakshina Kannada, Udupi and Kodagu districts and is a Council (NAAC) accredited four-star level institution.



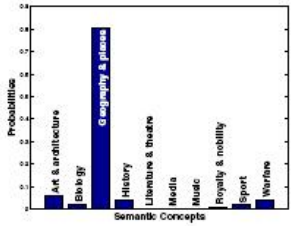
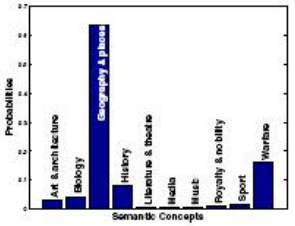
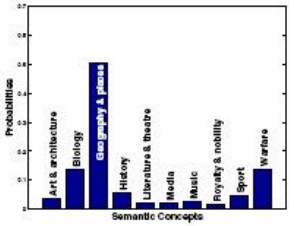
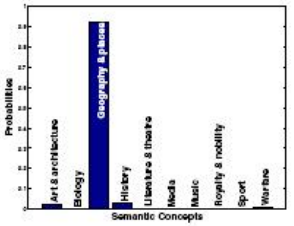
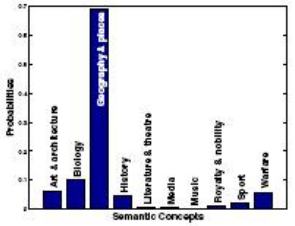
Text to image Retrieval

Query Text and its semantic vector

Retrieved Images and their semantic vectors





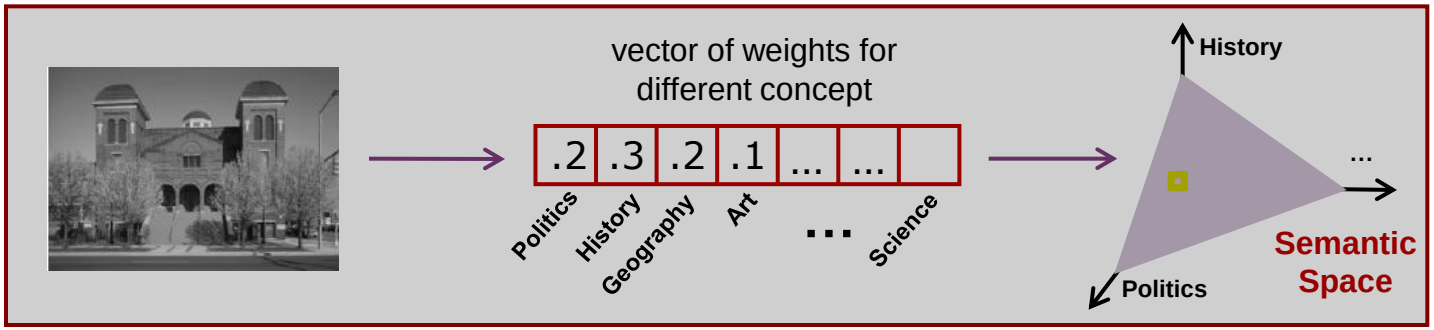
Comparison to unimodal retrieval

- Modality proxying

Experiment	Model	MAP
CCA + SMN (retrieved images proxy texts)	Log. Regression	0.277
CCA + SMN (image proxies text query)	Log. Regression	0.226
Image SMN [1]	Gauss Mixture	0.161
Image SMN	Log. Regression	0.152
Image SIFT features [2]	Gauss Mixture	0.135
Image SIFT features	Histogram	0.140
Random	-	0.118

Semantic Matching (SM)

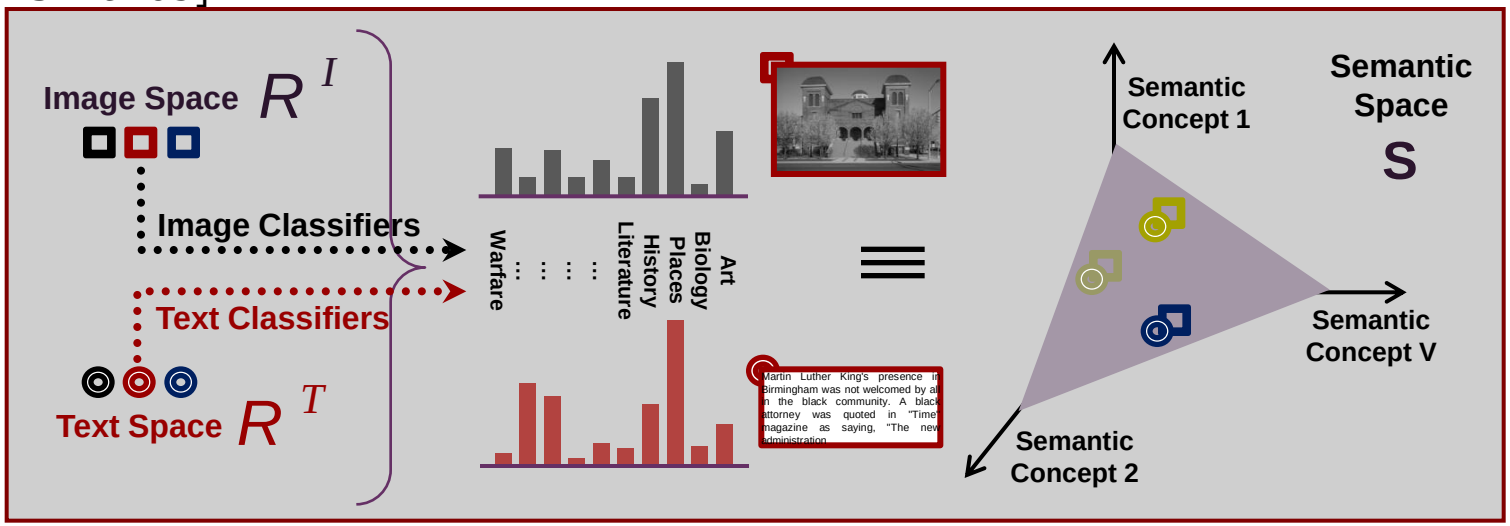
- Builds on the idea of Semantic Abstraction
  - transformation of data from the low-level feature representation to a high-level "semantic space" of meaningful concepts.



- To obtain Semantic Representation:-
  - Augment database with a vocabulary of semantic concepts
  - Learn classifiers for each concept
  - Represent each modality by its posterior probability under these learned classifiers
  - Thus the space is a "semantic simplex"

Semantic Matching (SM)

- Design semantic spaces for both modalities [Rasiwasia'07, Smith'03]



- We use multiclass logistic regression to classify both text and images
- The posteri… …iers serves as the semantic re…

$$\Pr(y_i = j \mid X_i) = \frac{\exp(X_i \beta_j)}{1 + \sum_{k=1}^{J} \exp(X_i \beta_k)}$$

Text/Image features

Learned parameters

Total number of classes

Acknowledgements

- Joint work with researchers from University of California, San Diego
  - Jose M. Costa Pereira,
  - Emanuele Coviello,
  - Gabriel Doyle,
  - Gert Lanckriet,
  - Roger Levy,
  - Nuno Vasconcelos

Questions?

© Bill Watterson

Training Phase: Step1 and Step 2

Compute Image Signatures

Retrieval/ Learning Operation

Image Pre-processing

+ Intensity
RGB
SIFT
SURF
HOG
BRISK
FREAK
…etc.

+ Direct
Bag-of-Words
TFIDF
SPM
VLAD
Fischer Vector
…etc.

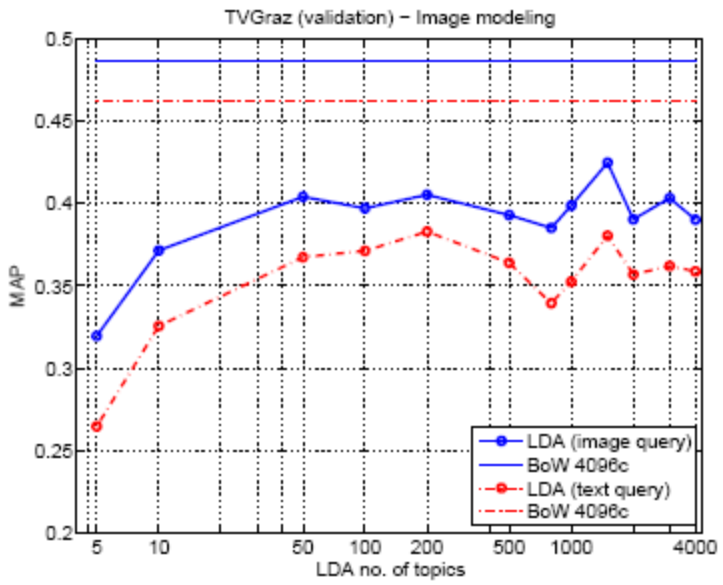
+ NN-Similarity
{Cosine, L1, L2, EMD, KL etc.}
Gaussian Mixture
SVM
Boosting
…etc.

Questions?

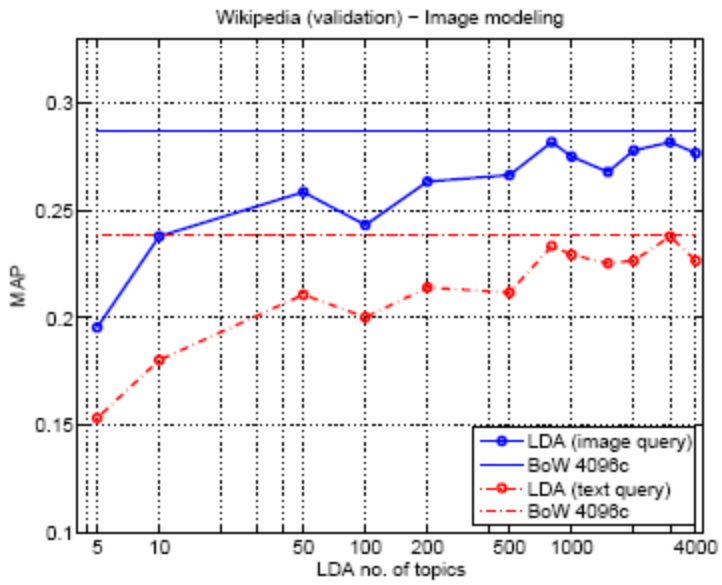
© Bill Watterson

Dimensionality Reduction for Images

- LDA can also be used to reduce image dimensions



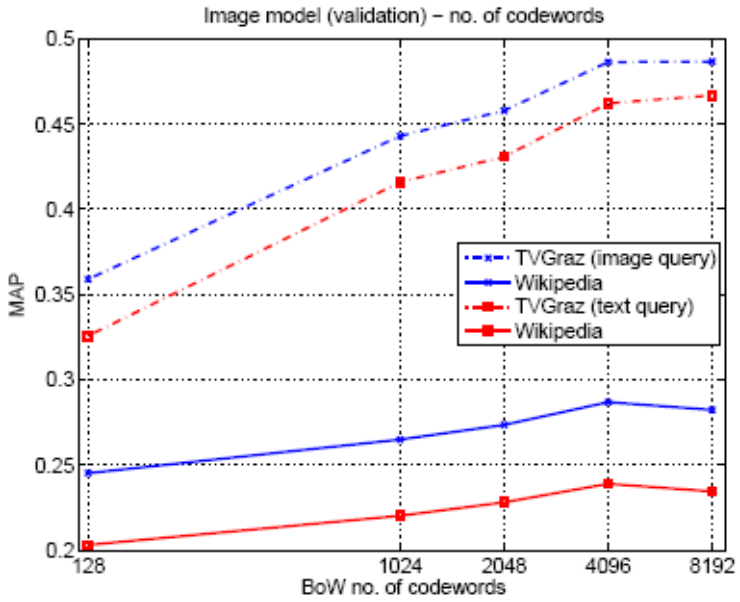
(a) TVGraz



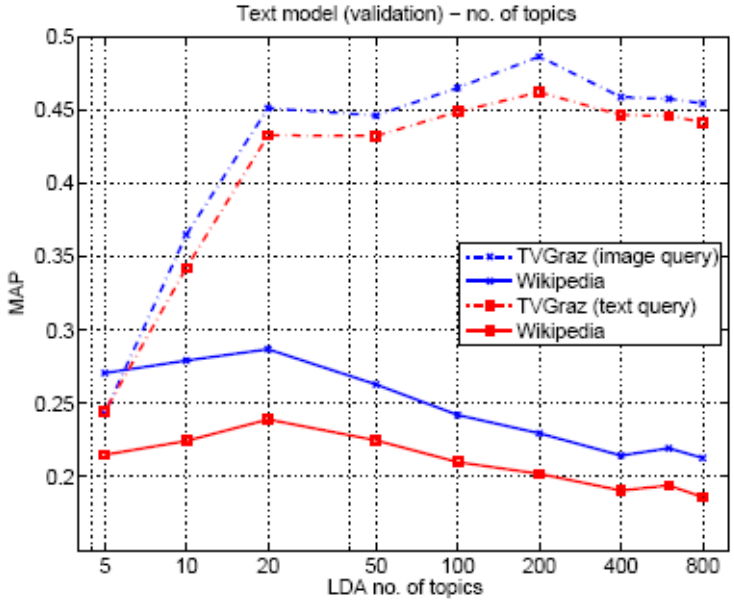
(b) Wikipedia

Parameter Validation

- Number of Image Codewords
- Number of Text Topics



(a) no. of codewords



(b) no. of topics

Distance Function

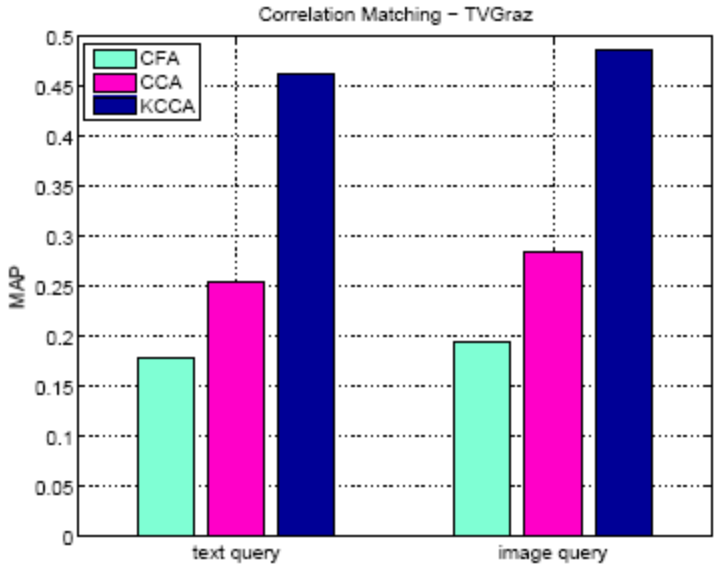
- Choice of distance function
  - For CM experiments - L1, L2, NC, NC_C (*)
  - For SM experiments - KL, L1, L2, NC, NC_C(*)

		TVGraz			Wikipedia		
Experiment	metric	img query	txt query	avg	img query	txt query	avg
CM	$L1$	0.376	0.418	0.397	0.193	0.234	0.214
	$L2$	0.391	0.444	0.417	0.199	0.243	0.221
	NC	0.498	0.476	**0.487**	0.288	0.239	**0.263**
	NC_c	0.486	0.462	0.474	0.287	0.239	**0.263**
SM	KL	0.296	0.546	0.421	0.188	0.276	0.232
	$L1$	0.412	0.548	0.480	0.232	0.276	0.254
	$L2$	0.380	0.550	0.465	0.211	0.278	0.245
	NC	0.533	0.560	0.546	0.315	0.278	0.296
	NC_c	0.579	0.556	**0.568**	0.354	0.272	**0.313**

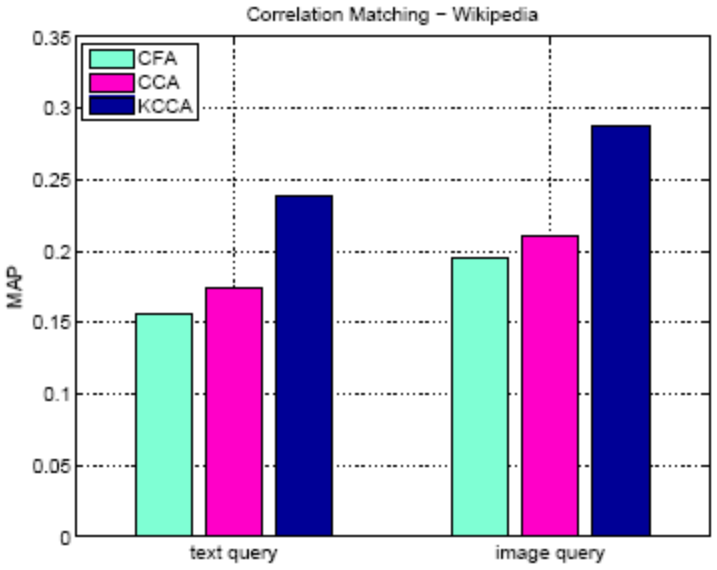
(*) $NC_c(p,q) = \frac{(p-\mu_p)^T (q-\mu_q)}{||p-\mu_p|| \, ||q-\mu_q||}$, where μ_p and μ_q are the mean for p and q.

Correlation Matching Methods

- KCCA vs CCA vs CFA



(a) TVGraz



(b) Wikipedia