Wrocław University of Technology

# PATSI – Photo Annotation through Similar Images with Annotation Length Optimization


Michał Stanek
Institute of Informatics


24-03-2011

# Presentation Outline

Automatic image annotation

# Presentation Outline

# Presentation Outline

# Presentation Outline

# Automatic Image Annotation

## Definition

### Image Annotation

Set of words from semantic dictionary $\mathcal{W}$ associated with an image $\mathcal{I}$

blue cloud outside plane sky

### Automatic Image Annotation

Annotator $\mathcal{A}$ describes previously unseen image $\mathcal{I}$ by a set of concepts $\mathrm{W}^{\mathcal{I}}$ from the semantic dictionary $\mathcal{W}$ based on the training dataset D, containing image–words pairs $(\mathrm{D} = \{(\mathcal{I}_1, \mathrm{W}^{\mathcal{I}_1}), \cdots, (\mathcal{I}_\mathrm{M}, \mathrm{W}^{\mathcal{I}_\mathrm{M}}))\})$.

# Automatic Image Annotation

- Annotation is a bridge between textual queries and visual content
- Find correlation between low-level visual features and high-level semantic
- Can be treated as multi-class classification problem (the number of classes is usually very large)
- Training data is often weakly annotated
  - annotation are incomplete,
  - may contain errors,
  - lack of association between the concept and the image region

# Example

## Treining dataset

Training Dataset

blue cloud outside plane sky

blue forest green outside road tree

# Example

## Query Image

### Query Image


??? ??? ??? ???

Questions:

1. what words should be assign to query image
2. how long should be the annotation

# PATSI – Photo Annotation through Similar images

## Hypothesis

Images similar in appearance are likely to share the same annotation

# Image representation

## Visual Features

- $\mathcal{I}$ – n-dimensional vector of visual features

$$v^{\mathcal{I}} = (v_1^{\mathcal{I}}, \cdots, v_n^{\mathcal{I}}) \tag{2}$$

- visual features are a m-dimensional vector of low level attributes

$$v_i^{\mathcal{I}} = (x_1^{i,\mathcal{I}}, \cdots, x_m^{i,\mathcal{I}}) \tag{3}$$

- low level attributes:
  - color
  - texture
  - edges
  - shape

# Minkowski Measure

## Distance in vector space

The Minkowski metrics between images $\mathcal{A}$ and $\mathcal{B}$ is defined as:

$$\mathrm{d}_{\mathrm{MK}}(\mathcal{A},\mathcal{B}) = \left(\sum_{\mathrm{i}=1}^{\mathrm{n}} \left|\mathrm{v}_\mathrm{i}^{\mathrm{A}} - \mathrm{v}_\mathrm{i}^{\mathrm{B}}\right|^{\mathrm{p}}\right)^{1/\mathrm{p}} \tag{4}$$

where p is the factor for the norm.
Particularly, when p is equal to one or to two, it is the well known L1 and Euclidean distance respectively.

# Jensen–Shannon Divergence

## Image model

Image model – Multivariate Gaussian Distribution

$$M^{\mathcal{I}}(x, \mu, \Sigma) = \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^{\top}\Sigma^{-1}(x-\mu)\right) \quad (5)$$

# Jensen–Shannon Divergence

## Information divergence

Jensen–Shannon Divergence between image $\mathcal{A}$ and $\mathcal{B}$ is defined as:

$$\mathrm{d}_{\mathrm{JS}}(\mathcal{A},\mathcal{B}) = \frac{1}{2}\mathrm{D}_{\mathrm{KL}}(\mathrm{M}^{\mathcal{A}} \| \mathrm{M}^{\mathcal{B}}) + \frac{1}{2}\mathrm{D}_{\mathrm{KL}}(\mathrm{M}^{\mathcal{B}} \| \mathrm{M}^{\mathcal{A}}), \tag{6}$$

Kullback–Leibler divergence:

$$\begin{aligned}\mathrm{D}_{\mathrm{KL}}(\mathcal{M}^{\mathrm{A}} \| \mathcal{M}^{\mathrm{B}}) &= \frac{1}{2}\log_{\mathrm{e}}\left(\frac{\det\Sigma_{\mathrm{B}}}{\det\Sigma_{\mathrm{A}}}\right) + \frac{1}{2}\mathrm{tr}\left(\Sigma_{\mathrm{B}}^{-1}\Sigma_{\mathrm{A}}\right) \\ &+ \frac{1}{2}\left(\mu_{\mathrm{B}} - \mu_{\mathrm{A}}\right)^{\top}\Sigma_{\mathrm{B}}^{-1}(\mu_{\mathrm{B}} - \mu_{\mathrm{A}}) - \frac{\mathrm{N}}{2},\end{aligned}$$

where $\Sigma_{\mathrm{A}}$, $\Sigma_{\mathrm{B}}$ and $\mu_{\mathrm{A}}$, $\mu_{\mathrm{B}}$ are covariance matrices and mean vectors from the respective image models A and B.

# Annotation Transfer

## Algorithm – Preparation Phase

### Preparation phase

1. Each image in a training database is divided into disjoint regions.
2. For all regions statistical visual features are calculated.
3. From the visual features create image model.

# Annotation Transfer

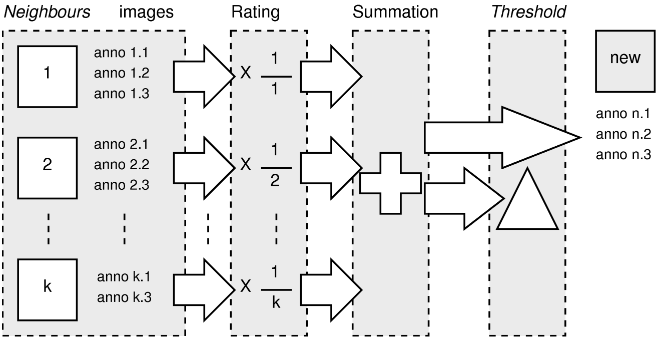
## Algorithm – Query Phase

### Query phase

1. Divide a query image into disjoint regions and calculate visual features vector.
2. Build the query image model from its visual features.
3. Calculate distance to all images models in training database
4. Take K images with smallest distances between models and create ranking of those images.
5. Transfer all words from images in the ranking with the value $\varphi(\mathrm{r})$, where r is the position of the image in the ranking, and $\varphi$ is a transfer function
6. As a final annotation take words which sum of the transfer values are greater or equal to provided threshold value t.

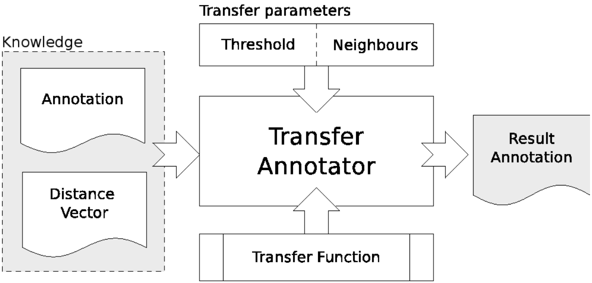
# Annotation Transfer

## Idea



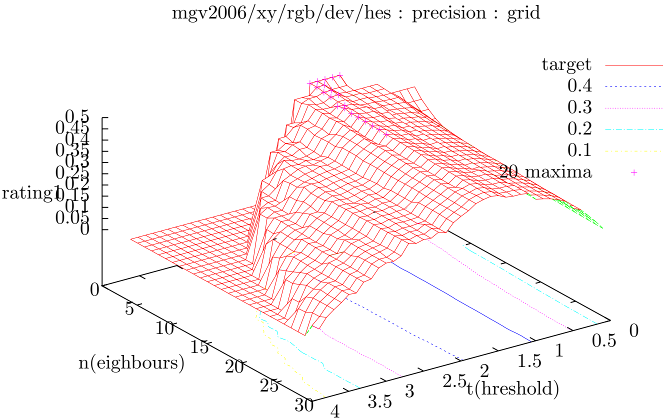
# Annotation Transfer

## Parameters



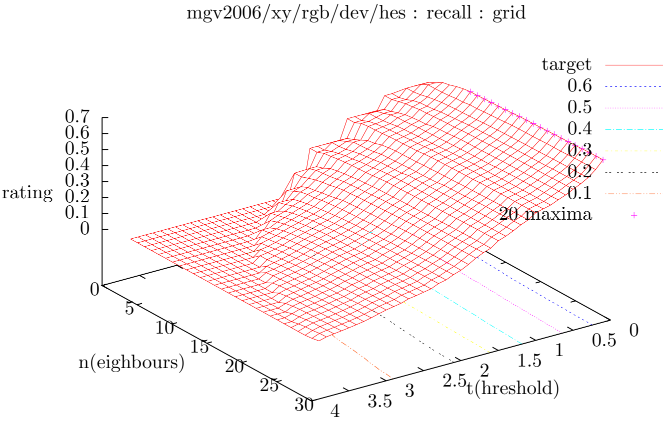
# Evaluation Measures

## Precision



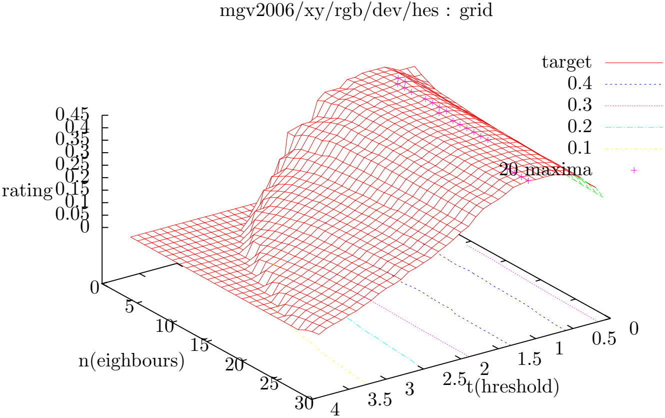
# Evaluation Measures

## Recall



# Evaluation Measures

## F-Score



# Transfer Parameters Optimization

## Problem definition

### PATSI Annotator

$$\mathcal{A}_{\mathrm{t,k}}(\mathcal{I}|\mathrm{d}, \varphi, \mathrm{D}), \tag{7}$$

where:

$\varphi$ – transfer function,

D – training dataset,

d – distance measure

# Transfer Parameter Optimization

## Optimization Criterion

### Optimization criterion

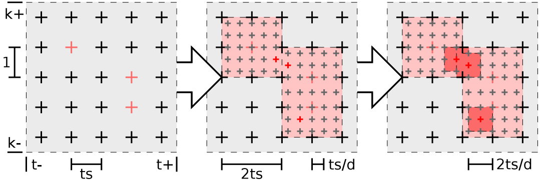
$$\phi(\mathrm{t},\mathrm{k}) = \frac{\sum_{(\mathcal{I},\mathcal{W}^{\mathcal{I}})\in \mathrm{D}} \mathrm{Q}(\mathcal{A}_{\mathrm{t},\mathrm{k}}(\mathcal{I}|\mathrm{d},\varphi,\mathrm{D}),\mathrm{W}^{\mathcal{I}})}{|\mathrm{D}|} \tag{8}$$

The optimal parameter values are therefore defined by

$$(\mathrm{t}^*,\mathrm{k}^*) = \underset{\mathrm{t},\mathrm{k}}{\mathrm{argmax}}(\phi(\mathrm{t},\mathrm{k})) \tag{9}$$

# Parameter Optimization

## Iterative refinement – Idea



# Benchmark datasets

## Statistics

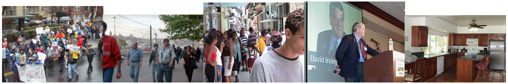
| | MGV 2006 | ICPR 2004 | IAPR TC-12 |
|---|---|---|---|
| Number of images | 751 | 1 109 | 19 805 |
| Dictionary size | 74 | 407 | 291 |
| Mean annotation length | 5.0 | 5.79 | 5.72 |
| Mediana of annotation length | 5.0 | 5.0 | 5.0 |
| Std. dev. of annotation length | 1.28 | 3.48 | 2.56 |
| Min. and max. annotation length | (2, 9) | (1,23) | (1,23) |

# Feature Sets

1. Mean values of H, S and V (in HSV colour space) and their std. deviations
2. Mean values of R, G, and B (in RGB colour space) and their std. deviations
3. Normalized X and Y coordinates of center of the region, mean R, G, and B (in RGB colour space), std. deviation of R, G, B and mean eigenvalue of the colour Hessian computed in RGB colour space

# Most similar images to query image

Table: Similar images using JS divergence and 3rd feature set.

| Query image | Similar images |
|---|---|
|  |  |
|  |   |
|  |   |

# Results

## MGV 2006 dataset

| Method | Precision | Recall | F-Score |
|---|---|---|---|
| FastDIM | 0.24 | 0.16 | 0.19 |
| FastDIM + GRWCO | 0.34 | 0.34 | 0.34 |
| MCML | 0.32 | 0.24 | 0.27 |
| MCML + GRWCO | 0.38 | 0.37 | 0.37 |
| CRM | 0.39 | 0.34 | 0.36 |
| PATSI | 0.38 | 0.46 | 0.42 |
| The best 20 words | | | |
| FastDIM | 0.58 | 0.53 | 0.51 |
| FastDIM + GRWCO | 0.59 | 0.61 | 0.60 |
| MCML | 0.61 | 0.59 | 0.60 |
| MCML + GRWCO | 0.64 | 0.62 | 0.63 |
| CRM | 0.58 | 0.57 | 0.57 |
| PATSI | 0.71 | 0.86 | 0.78 |

# Results

## ICPR 2004 datasets

| Method | Precision | Recall | F-Score |
|---|---|---|---|
| FastDIM | 0.20 | 0.17 | 0.18 |
| FastDIM + GRWCO | 0.21 | 0.21 | 0.21 |
| MCML | 0.21 | 0.17 | 0.19 |
| MCML + GRWCO | 0.25 | 0.28 | 0.26 |
| CRM | 0.24 | 0.24 | 0.24 |
| PATSI | 0.27 | 0.34 | 0.30 |
| The best 60 words | | | |
| FastDIM | 0.64 | 0.58 | 0.61 |
| FastDIM + GRWCO | 0.63 | 0.61 | 0.62 |
| MCML | 0.69 | 0.60 | 0.64 |
| MCML + GRWCO | 0.69 | 0.67 | 0.68 |
| CRM | 0.61 | 0.61 | 0.61 |
| PATSI | 0.82 | 0.94 | 0.88 |

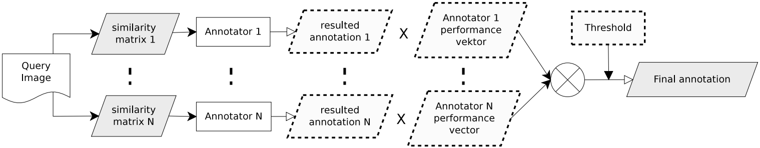
# Results

## IAPR TC 12

| Method | Precision | Recall | F-Score |
|---|---|---|---|
| RGB | 0.24 | 0.24 | 0.24 |
| HSV | 0.20 | 0.20 | 0.20 |
| LAB | 0.24 | 0.25 | 0.24 |
| Haar | 0.20 | 0.11 | 0.14 |
| HaarQ | 0.19 | 0.16 | 0.17 |
| Gabor | 0.15 | 0.15 | 0.15 |
| GaborQ | 0.08 | 0.09 | 0.08 |
| MBRM | 0.24 | 0.23 | 0.23 |
| Lasso | 0.28 | 0.29 | 0.28 |
| JEC | 0.28 | 0.29 | 0.28 |
| PATSI | 0.26 | 0.31 | 0.28 |

# PATSI Extensions

Use many features and multiple distance measures




Possible strategies:

- best – for each word take the best distance measure
- weighted – take many distance measures and weight them according to performance

# Multiple distance measures

| Method | Precision | Recall | F-Score |
|---|---|---|---|
| FastDIM | 0.24 | 0.16 | 0.19 |
| FastDIM + GRWCO | 0.34 | 0.34 | 0.34 |
| MCML | 0.32 | 0.24 | 0.27 |
| MCML + GRWCO | 0.38 | 0.37 | 0.37 |
| CRM | 0.39 | 0.34 | 0.36 |
| PATSI | 0.38 | 0.46 | 0.42 |
| multiPATSI – best | 0.44 | 0.65 | 0.53 |
| multiPATSI – weighted | 0.58 | 0.66 | 0.62 |

# Conclusion

## PATSI

- is simple – could be used as new baseline for image annotation
- could determine the resulted annotation length
- achieve very high recall on training datasets
- preserves annotation words coocurances distribution