

Slide Image Retrieval: A Preliminary Study

Guo Min Liew and Min-Yen Kan

National University of Singapore
Web IR / NLP Group (WING)

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture

Beyond Text in Digital Libraries

Papers don't store all of the information about a discovery

- Dataset
- Tools
- Implementation details / conditions



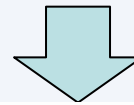
We'll focus on this

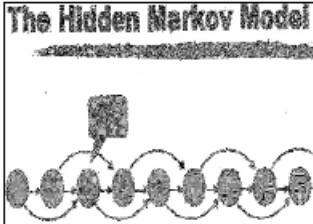
They also don't help a person learn the research

- Textbooks
- Slides



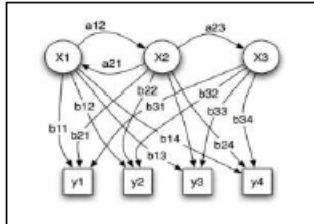
What is Slide Image Retrieval?





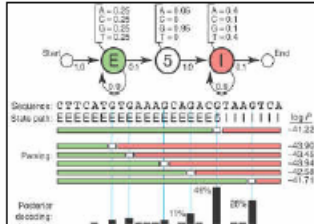
1. Hidden Markov Model

Slide Number: 5



2. Chromosome...Hidden Markov Models

Slide Number: 14



3. Chromosome...Hidden Markov Models

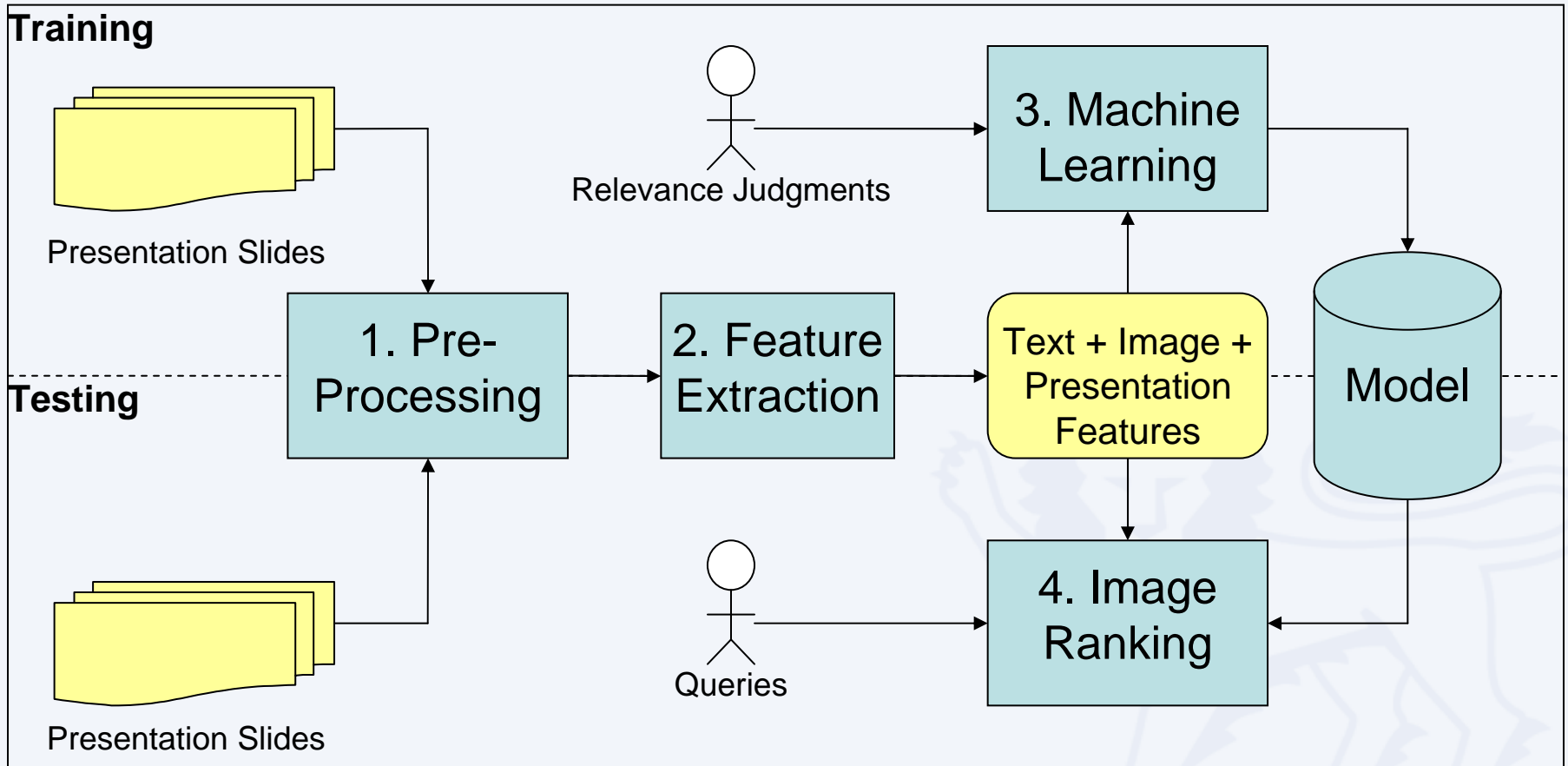
Slide Number: 23

- Hypothesis: **image** and **presentation** features can enhance text-based features

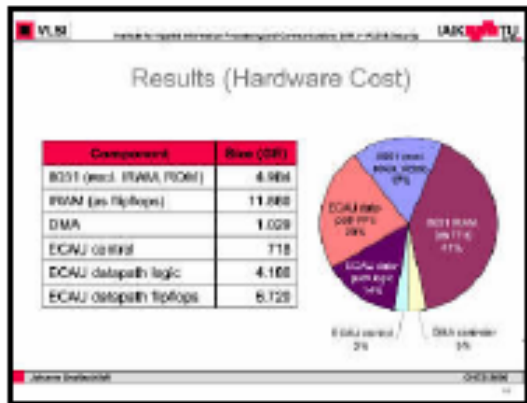




Slidir Architecture



1. Preprocessing



Presentation Slide

Simple preprocessing, not meant to be state-of-the-art

Generally consistent background → Perform collective separation

Top down approach using projection profile cuts with thresholds from PPT template

Sample Extraction Results

(a) Original Slide

Motivation

- Problem**
 - ECC is very computation-intensive
 - 8-bit micro-controllers like 8051 are too slow for ECC
- Goal**
 - ECCSA on 8051 in less than 1 sec (scalar mult. in $O(5 \text{ sec})$)
 - Low-cost hardware accelerator (minimal area)
- Hardware/software co-design**
 - Field arithmetic in HW, rest in SW
 - 192-bit Arithmetic Unit (approx. 11k gates)
 - Addition and Multiplication in GF(2^m)
 - Scalar multiplication over GF(2^m) in 120 msec @ 12 MHz

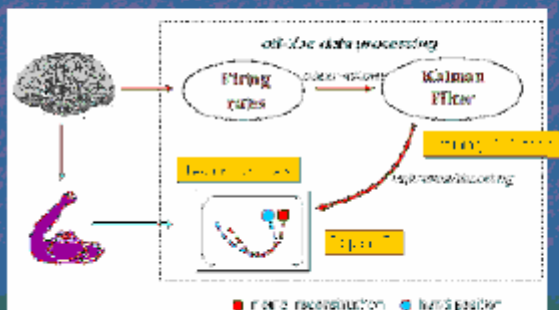
(b) Extracted Background

(c) Extracted Foreground

Motivation

- Problem**
 - ECC is very computation-intensive
 - 8-bit micro-controllers like 8051 are too slow for ECC
- Goal**
 - ECCSA on 8051 in less than 1 sec (scalar mult. in $O(5 \text{ sec})$)
 - Low-cost hardware accelerator (minimal area)
- Hardware/software co-design**
 - Field arithmetic in HW, rest in SW
 - 192-bit Arithmetic Unit (approx. 11k gates)
 - Addition and Multiplication in GF(2^m)
 - Scalar multiplication over GF(2^m) in 120 msec @ 12 MHz

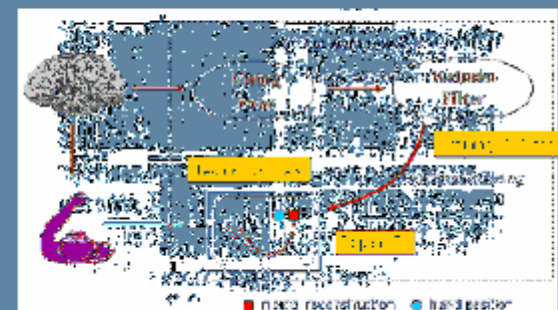
Experimental Paradigm



Background



Experimental Paradigm



2. Feature Extraction

Utilizes **textual**, **image** and **presentation** features

- **Textual** Features

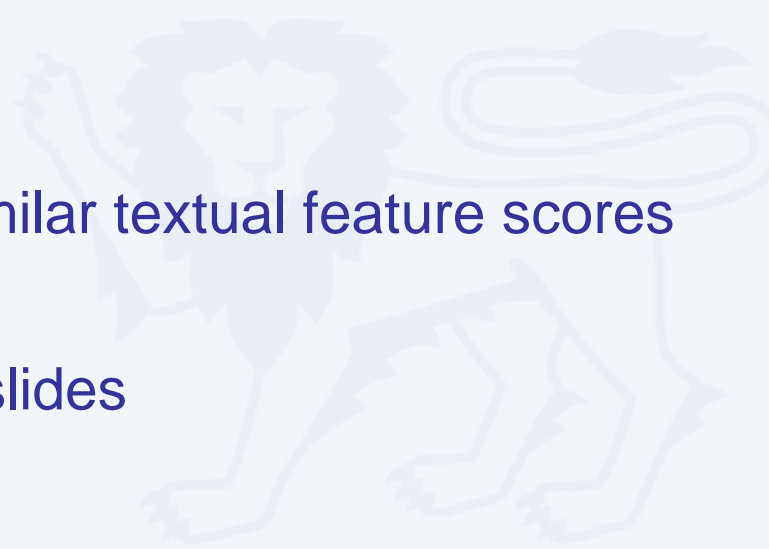
- Main component of feature set
- Most images identified by surrounding and embedded text

- **Image** Features

- Supplement to textual features
- Useful in ranking images with similar textual feature scores

- **Presentation** Features

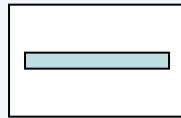
- Features unique to presentation slides



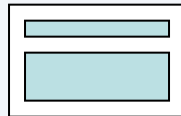
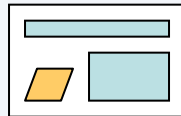
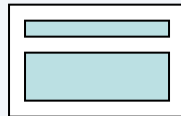
Sample Features

Textual

1. All text
2. Slide text
3. Next slide's text
4. Previous slide's text
5. Slide Title
6. Presentation Title
7. Slide Image Text

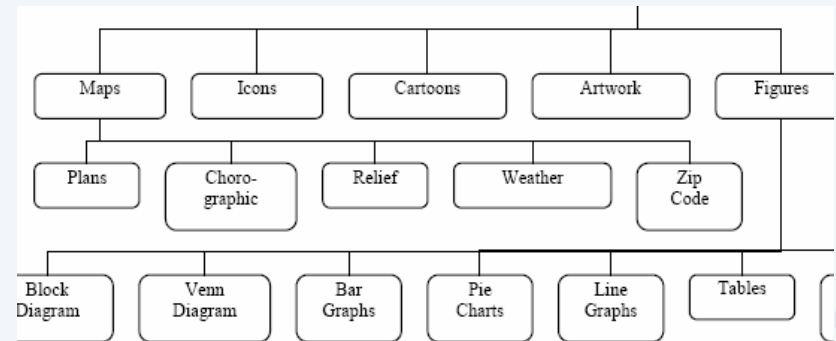


⋮



Image

- Image Size
- Image Colors
- NPIC image type (CIVR, 06)



Presentation

1. Slide Order
2. Image position on Slide
3. Number of Images on Slide

Machine Learning

Regression Model

- Relevance judgments and feature vectors input into machine learner
- Linear regression used to produce continuous output value for image ranking

Human Relevance Judgments

- 9 participants
- Four presentation sets of 10 queries each
- Participants rank up to five most relevant images to the query

Regression Analysis

- 10-fold cross validation
- System with fielded text features able to replicate human upper bound
- Image and presentation features further improve correlation

	All text baseline	Fielded Text	With Presentation and Image Features	Inter-annotator
Mean absolute error (lower is better)	19.36	13.72	13.81	9.40
Correlation (higher is better)	-0.015	0.565	0.581	0.531

Still a gap to improve

Looks pretty good

Comparison of mean absolute error and correlation

Binary Classification

- Human participants judgments converted to binary
- Evaluated using SVM and J48 decision tree classifier
- General accuracy fairly good
- Decision tree algorithm slightly more precise → sequential testing adequate

	SVM (SMO)	J48 Decision Tree
Accuracy	90.9%	92.5%
Kappa	.54	.59
Relevant Class Precision	.59	.70
Recall	.58	.58
F₁	.59	.63


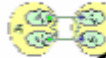
10-fold performance over 8062 instances

Feature Analysis

- Human subject study with 20 volunteers
- Subjects select most and least relevant images

Does This Idea Work?

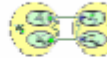
- Check whether a graph is valid or not
- Find the least number of vertices to delete from a graph
- Find the number of vertices to delete from a graph
- The number of vertices to delete

A B

Odd Hole / Two Property 2-Join Decom



- Minimum number of vertices to delete from a graph to make it bipartite
- Minimum number of vertices to delete from a graph to make it bipartite
- Minimum number of vertices to delete from a graph to make it bipartite
- Minimum number of vertices to delete from a graph to make it bipartite



C

Odd Hole / Two Property 1-Vertex Clear Graph


- Minimum number of vertices to delete from a graph to make it bipartite
- Minimum number of vertices to delete from a graph to make it bipartite
- Minimum number of vertices to delete from a graph to make it bipartite
- Minimum number of vertices to delete from a graph to make it bipartite

D E

Idea of Overcoming the Problem State of Infeasible Problem

- Minimum number of vertices to delete from a graph to make it bipartite
- Minimum number of vertices to delete from a graph to make it bipartite
- Minimum number of vertices to delete from a graph to make it bipartite
- Minimum number of vertices to delete from a graph to make it bipartite



F

Sample Human Subject Survey

Presentation Set 3							
For each query, choose the most relevant images from the slides and rank them accordingly. Stop if no more relevant images (see sample)							
No.	Query	Most Relevant Image by Rank					Comments
		1	2	3	4	5	
1	Perfect Graphs	A	C	E	B	na	
2	Odd Hole						
3	2-join						
4	double star						
5	Xinning Liu						

Rank	SLIDIR Features	Regression Coefficients
1	Slide Title	-234.39
2	Slide Image Type	-124.09
3	Slide Text	-83.86
4	Slide Image Text	-83.64
5	Next Slide's Text	-29.42
6	Presentation Title	-9.85
7	Number of Images on Slide	-1.89
8	Slide Order	-0.674
9	Image Size	0.0
10	Slide Image Position	0.0
11	Number of Colors	0.0004
12	Previous Slide's Text	31.70

Textual features most important, **image** and **presentation** features still play significant role

Image Ranking Evaluation

- Online survey with 15 volunteers
- 8 given queries and sampled dataset of 100 images
- Volunteers indicate relevant images to query
- 53% of top 10 images are relevant (max of 15 relevant)
- Average precision of 0.74

	Precision at n	Recall at n
n = top 10	.53	.73
n = top 20	.38	.95
n = top 30	.27	.98
n = top 40	.21	1.0
Average Precision	.74	

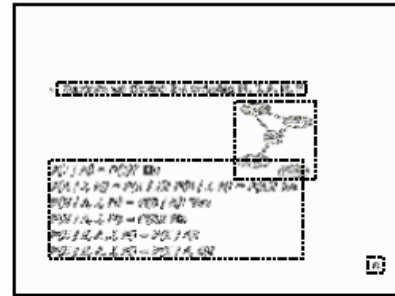
Future Work

Limitations

- Very small data set – 100 images from an AI course

Preprocessing

- Improve segmentation and collective extraction



Field Implementation – Ongoing work

- Incorporate with previous work on SlideSeer (JCDDL, 2007) that aligns presentation and document pairs

More information:

Liew Guo Min(2008) **SLIDIR: Slide Image Retrieval**, Undergraduate Thesis, School of Computing, National University of Singapore.



See you in Singapore!
Next month or next year!

Questions?

Welcome to ACL-IJCNLP 2009!

A Joint Conference of
the 47th Annual Meeting of the Association for Computational Linguistics and
the 4th International Joint Conference on Natural Language Processing

Suntec Convention Centre • Singapore

August 2-7, 2009

www.acl-ijcnlp-2009.org

Local Organiser

