Towards Web-Scale Media Content Analysis & Retrieval:
What has University Research Contributed to Commercial Systems and Social Network Services

Tat-Seng Chua
Lab for Media Search, School of Computing
National University of Singapore
5 July 2010
Outline of Talk

- Information Rich World
- Contributions of MM Research
- Achievements of MM Research
- Bridge the Semantic Gaps
- Summary
Rich Multimedia Contents

• MM contents are growing at exponential rate

• Flickr
  o Receives over 3,000 photos upload every minute, or >4 million per day
  o Over 3 millions with geo-tagged each month

• YouTube
  o Statistics of Uploads:
    ▪ 78.3 million up-loaded as of March 2008
    ▪ > 200,000 per day
    ▪ 10 hours of video every minute
  o Statistics of Downloads:
    ▪ >100 millions per day
    ▪ Over 60% of video watched on Internet
    ▪ Over 20% of Web traffic is for YouTube video
Popular Social Network Sites

• Even text-based sites contains lots of images

• **Wikipedia:**
  o Has over 1 million topics, most with media contents
  o Has over 1 million geo-tagged articles

• **Yahoo!Answers:**
  o Over 15% of answers now have links to other sites, mostly for media contents

• **Facebook:**
  o Has huge volume of photos, though tagging is sparse and less content-oriented

• **Twitter:**
  o Many posting are with links to other media sources

• The field is vibrant; we can find almost any contents on the Web
What About Locations

- After Google Earth, Google Map, What’s next?
- Yahoo!WOEID uniquely reference spatial entities -- from country to suburb
- **FourSquare**: Encourage users to tag & check-in places in Game like environment
  - Users check-in places to gain power and incentives
  - Others can find places and whereabouts of friends
  - Over 1.8 millions users
    - growth rate faster than Twitter at similar stage
Characteristics of Successful Commercial Ventures

- Fun / Consumer oriented
  - Act on principles of utilizing (stealing?) your idle time
  - Accuracy is never the criterion
- Serve social functions
- Identify items based on context – not contents
  - Most are text-based (tag or concept-based)
- Focus on user experience – simplicity is the key
Outline of Talk

- Information Rich World
- Contributions of MM Research
- Achievements of MM Research
- Bridge the Semantic Gaps
- Summary
Contributions of MM Research?

- Has years of MM /vision research contributed towards success of these ventures?

- One (or majority) point of views:
  - Very little
  - Most big ideas are simple !!
  - We are benefiting form their success, rather than helping them
    We do research on their data (Flickr, YouTube, Twitter etc)..

- Reasons:
  1) They are not technology oriented – put up brave face
  2) Other CS research has not contributed much either – lay blames on all
  3) They are consumer oriented where accuracy is not important, and context info is more than sufficient – is only context sufficient?
  4) No real data for large-scale academic research – only toy problems
Contributions of MM Research

- Context-only approach has its limitation
  - As such companies evolved, media content analysis becomes more important

- Not all glooms and dooms, some success stories
  - Duplicates removal in YouTube
  - Large-scale visual matching in Bing/Google Image Search
  - Landmark matching
  - Snap-Tell applications.
  - Others..
Content-based Search in Commercial Search Engine

- For example, Google Image Search, Bing etc.
A Reverse Image Search Engine

- It finds out where an image came from, how it is being used
- Copy and Near duplicate detection technology?
- Purely content-based search on database of over 1.5 billion images
Many similar examples in the market!!
Mostly based on fast image matching

Examples of Successful Matches

Effective Image Matching Algorithm + Unprecedented Scale
- Mobile phone app: identify works of arts thru photos
- Acquired by Google
- Similar Snap&Tell mobile app
- Match object in database on millions of objects in Amazon’s EC2 Cloud Computing Services
Image Labeling

Mobile Visual Search
- List of objects searchable via Goggle
  - Snap pictures and get answers

- More details of this Landmarks

- Give me the English translation of this menu?

- Key technologies: OCR, logo detection, image matching…
Outline of Talk

- Information Rich World
- Contributions of MM Research
- Achievements of MM Research
- Bridge the Semantic Gaps
- Summary
Achievements on MM Research

- Most current achievements centered on following technologies
  - Duplicate /copy detection
  - kNN-based fast image matching
  - Visual concept annotation
  - Indexing
  - OCR
  - Landmark recognition
Content-based Media Search- History

- **Use of media contents in search evolved in the early 90s**
- **QBIC of IBM** is an early example (Flickner et al ’95)
  - First **complete commercial CBIR System** that supports Query by Image Content
  - Uses color, texture, shape features
  - Text-based search can also be combined
  - Uses R*-trees for indexing
More recently, content-based media search has been integrated into main stream search engines.

Why after more than 15 years since QBIC?

- Local visual features, such as SIFT, is sufficiently robust
- Large amount of social annotations available
- Advances in visual concept annotation and indexing strategy permit large-scale kNN visual search to be performed efficiently
- Other technologies like: Duplicate detection etc
Near Duplicate Detection Techniques

- **Signature-based:** *fast but not robust*
  (W. Dong et al. ACM MM 08, Chum et al. BMVC 08)

- **Keypoint matching:** *robust but slow*
  (Ke et al. ACM MM 2005, Ngo et al. ACM MM 06)

- **Trajectory-based:** *sensitive to camera motion*
  (Wu et al. ACM MM 08, Law-to et al. ACM MM 06)

- **Content + Meta data:** *real-time near-duplicate elimination*
  [Wu et al. TMM 09]
Partial Near-Duplicate Detection

- Hyper-linking videos
  - Non-overlapping parts bring in new information of a topic

- Further work by bringing temporal network to improve performance

- Applications
  - Redundant content removal
  - Content re-ranking
  - Content analysis
Visual Concept Annotation -1

• Given an image or video segment
  o Determine whether a visual concepts is present, e.g. face, car, road, fire etc.
  o Semantic Gap Problem

• The premise: user-supplied tags provide huge amount of free training data for Web-based image collections

• Key question is whether these social tags are good enough?
  o They are noisy & incomplete ⇒ need noise removal
  o Less than 10% of images in Flickr have tags ⇒ need concept propagation

• Hot topic of research – 2 main approaches adopted:
  o Traditional machine learning approach… but need training data
  o Utilizing the metadata in the web … but noise problem
  o No free-lunch!!!
Visual Concept Annotation -2

- Traditional Approaches:
  - Learning based: CMRM, SVM, LDA, k-NN, SSL,…
  - Active in TRECVID evaluations, like VIREO, DVMM, Media-Mill
  - Many have released hundreds of concept detectors, but performance are mostly limited

- Main drawbacks: needs large amount of manually annotated training data
Visual Concept Annotation - 3

- “80 million tiny images: a large dataset for non-parametric object and scene recognition” (Torralba et al. PAMI’ 08)
  - Features: vector of pixel values
  - Method: kNN with Hashing
  - Problem: effective only for some categories (with good variety of samples)
  - Essentially rely on fast matching of great variety of same kinds of images – brute force approach??
Visual Concept Annotation

- "Inferring from Flickr images with tag noise handling" (Tang et al. ACM MM’09)
  - Features: visual & text
  - Method: Sparse Graph Semi-Supervised Learning
  - Key focus: Noise removal in metadata, but has problem of efficiency

Tested on NUS-WIDE-Lite:
- 55,615 images
- SGSSL with tag noise handling performs the best with MAP = 0.16
Visual Concept Annotation

- “Large-Scale Multi-Label propagation” (Chen et al...)
  - Hashing-based L1-Graph Construction:
  - KL-Divergence-based Probabilistic Multi-label Propagation
    ▪ achieve robustness

Mean Average Precision: 0.215 on NUS-Wide dataset

Improvement of MAP from 0.16 to 0.215, does it make any difference to user’s search experience??
Web-Scale Media Indexing

• Task in Image Search
  ○ Find database images that are similar to query image

  ![Image Database](image_database.png)
  ![Query Image](query_image.png)
  ![Feature Extraction](feature_extraction.png)
  ![Visual feature space](visual_feature_space.png)
  ![Indexing](indexing.png)
  ![Search](search.png)
  ![Index database](index_database.png)

• Challenges
  ○ Large-scale image database
  ○ High-dimensional visual feature
Web-Scale Media Indexing - 2

- **Spatial Indexing:**
  - Divide the feature space based on the distribution of images
  - HB-tree [Lomet, 90], X-tree [Berchtold, 96], M-tree [Ciaccia, 97], etc.

- **Hash-based Indexing:**
  - Hash the similar images into the same bucket based on a family of hash functions
  - LSH [Ciaccia, 97], Semantic Hashing [Salakhutdinov, 08], Spectral Hashing [Weiss, 09], etc.

- **Others:**
  - Data Compression, like VA-file [Weber, 98], VQ-Index [Tuncel, 02], etc.
  - Dimension reduction and hybrid of tree-like structures
  - etc.
More Semantic-oriented Hashing

• Limitation of current hashing approaches
  - Only preserve visual proximities between images in Hamming space
  - May return unsatisfactory search results due to Semantic Gap

• Metadata associated with images
  - Class labels
  - Tags
  - Relevance feedbacks
  - Click-through data

Towards new hashing approach

• Learn hash functions by simultaneously leveraging the visual content and the metadata associated with images
• Design hash functions that can index any type of data
• Recent works in U Columbia, NUS etc.
Aim: to build a Web-scale landmark recognition engine

“Tour the World: building a web-scale landmark recognition engine” (Zheng et al. CVPR 09)
Discovering Landmarks in the world

- Two approaches to discover landmarks:
  - Photos in photo sharing websites
  - Online tourist articles

- Discovered 5312 landmarks
  - 1,259 cities, 144 countries
Unsupervised learning of landmark images

Geo-clusters

Landmarks from tour articles

Noisy image pool

Visual clustering

Validate and clean models

Premise: photos from landmark should be similar

Clustering based on local features

Visual model validates landmarks!

Photo v.s. non-photo classifier to filter out noisy images
Visual Cluster Examples

- Retrieval experiments:
  - Index using k-d tree
  - Perform kNN style search on clusters
  - Achieve >80 accuracy on over 400 test images

- Visual Cluster examples:
  
  ![Visual Cluster Examples]

Acropolis, Athens, Greece

Corcovado, Rio de Janeiro, Brazil
Reasons for High Effectiveness

- **Web-scale and variety of examples**
  - For popular landmarks, we have seen most relevant examples from most viewpoints and illumination
  - Simple clustering permit organization of data for effective kNN style search
  - Similar conclusion from MIT experiment
    “80 million tiny images: a large dataset for non-parametric object and scene recognition” (Torralba et al. PAMI’ 08)
Outline of Talk

- Information Rich World
- Contributions of MM Research
- Achievements of MM Research
- Bridge the Semantic Gaps
- Summary
Towards Effective Web-Scale Media Processing

- How to further improve the quality and efficiency of Web-scale media search?
- Need to move from text-based towards **concept-based** processing
- Need to overcome **semantic gap**?
  - See dog example!!
  - Either with user in the loop (with min. efforts)
  - Or utilize **domain knowledge**

- Lots of research efforts..., but progress is slow.
Homan-in-the-Loop Solution

• ESP Game: a successful example of leveraging human efforts for image annotation
  o 2-player game – jointly tag an image – by guessing common keywords
  o 3.2 million labels with 22,000 users
  o Many people played over 20 hours a week

• Problem with ESP Game approach
  o Users tag one image at a time
  o Tedious and requires a huge no. of “players”

• Our aim:
  o Permit one person to manually annotate million images with multi-labels within a short time
  o **Follow ancient wisdom: adopt “万人敌”-type solution**

西楚霸王项羽: “学文不过能记住姓名，学武不过能以一抵百，籍要学便学万人敌！”
Fast Semi-Auto Labeling Framework - 1

• Basic idea:
  - Get user to label clusters, instead of individual images
  - But for effectiveness, cluster objects at region level using, say, LSH
  - But it is often difficult to recognize identity of regions, hence need to present region as part of image (the context)
Fast Semi-Auto Labeling Framework -2

- Basic idea (cont.):
  - Different situations to label: Most cluster belongs to a unique label; 2 or more labels, or background
  - User relevance judgments are feedback to refine multi-scale clustering
    - Need to perform a combination of label propagation, label difficulty estimation and uncertain image selection

![Diagram](image)
Auto Concept-Based Video Search

• Typical concept-based video search strategy
  o Concept Selection: Map text query to related concepts
  o Search with concepts: Retrieve videos using each concept as query
  o Fusion of search results: Generate final search results by fusing individual results from concepts

• Limitations
  o Difficulty in interpreting complex queries that are usually in the form of phrases or sentences (involving multiple concepts)
  o Limited concept detectors vs. unconstrained users’ queries
User-in-the-Loop Interactive Concept-based Video Search

- Auto search has achieved limited success
- Interactive search – TRECVID Experience
  - User labeling: Search engine presents results for users to label some as relevant and rest as irrelevant samples
  - Results updating: Search engine updates results based on users’ labeling
  - Lead to 3-fold improvement in accuracy

Main Challenge
- Insufficient relevant sample problem: Interactive search is ineffective if no relevance samples can be found
- Relevance feedback strategies
- Good user experience

Demo1, Demo2
Our Approach to: Sparse Relevant Sample Problem

• Utilizing Related Samples in RF process
  o Assumption: relevant samples are near related samples

• Utilize Related Samples in learning process:
  o Method arranges related samples in margin area (between the relevant and irrelevant samples)
  o Help to induce more relevant samples in subsequent iterations
  o More details in another talk
Location Search

- We are quite successful in searching for popular landmarks
  - Reason: we have most relevant examples from most viewpoints and illumination
  - Simple clustering permits organization of data for effective kNN style search

- What about the difficult problems?
  - Cases with few samples, like your neighbor locations?

- What can we do?
  - We can leverage on not just Web images, but also various knowledge sources & social location tagging info
Info from Knowledge Source: Yahoo! WOEID

- WOEIDs uniquely reference spatial entities provided by Yahoo!GeoPlanet
  - Coverage: ≈six million named places globally
  - Special Place Types: Postal Code; Super name; Colloquial names; Time Zone:, etc.

- Yahoo! GeoPlanet provides translation from WOEID to place names and vise versa.

- Yahoo! Placemaker determines what places are referenced in a unit of text by returning the WOEIDs and the names of the places found within a query document
Location Information

- With WOEID + Goggle Earth, we can map to town/suburb level
Location Information

- Town/suburb details + common places of interests

- But what about local information, like market, office, etc??
Local Location Info from Social Networks

- Encourages users to "check-in" at venues using a mobile website, text messaging or a device-specific application
- Users are then awarded points and sometimes badges, majorship, free goodies
- Search: Find list of nearby places; whereabouts of friends, etc
- Allows many local places to be tagged
Location Information

- Now plus local places from 4Square, we have...

- Combination of 3 sources of info provides rich wealth of location info...
Outline of Talk

- Information Rich World
- Contributions of MM Research
- Achievements of MM Research
- Bridge the Semantic Gaps
- Summary
Summary

- Current States:
  - Technologies in media search have evolved and matured to tackle a wide range of issues
  - Wide range of data, knowledge sources and social network data available online – for entities & locations
  - People are attracted to interesting apps that helps them to find /aggregate info at their “idle” time

- Research Opportunities:
  - Leveraging these info sources for interesting research and applications
  - Combine with QA offers interesting possibilities
  - Many practical problems with high impact
Q & A

Contact: Chua Tat-Seng (chuats@comp.nus.edu.sg)