



Multimedia Retrieval: *Are We Doing it Right?*

CHUA Tat-Seng

National University of Singapore

OUTLINE

- Retrospective on Image Search Since 2010
- Why Not Much Research in Video
- Some Current MM Research Efforts
- Summary

Disclaimer: The coverage of this talk is very broad, and I have time to cover only topics and works close to me, and do not have time to highlight interesting related works from you.

Two Basic Questions?

What is Multimedia
Retrieval?

Why from 2010?

Invited Talk at CIVR 2010



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

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..

Introduced Several Systems

Con
s



IQ Engines



Google Goggles

Use pictures to search the web.

- List of objects searchable via Google
 - Snap pictures and get answers



- More details of this Landmarks



- Key technologies: OCR, logo detection, image matching...

- Give me the English translation of this menu?



LMSearch

LMSearch

Technologies Behind These Early Commercial Systems?

- Robust image matching technologies
- Hashing for large scale image indexing
- Concept Annotation for fixed number of common categories

What happen to these early systems??



SnapTell

- Founded in 2006 in Silicon Valley; obtained \$4 million USD fund in Round 1
- It supports “snap and find” of covers of CD, DVD, book, or video game; returns product with ratings and pricing info from Google, Amazon, eBay and more.
- Acquired by Amazon on June 16, 2009
To add technology to its mobile offering



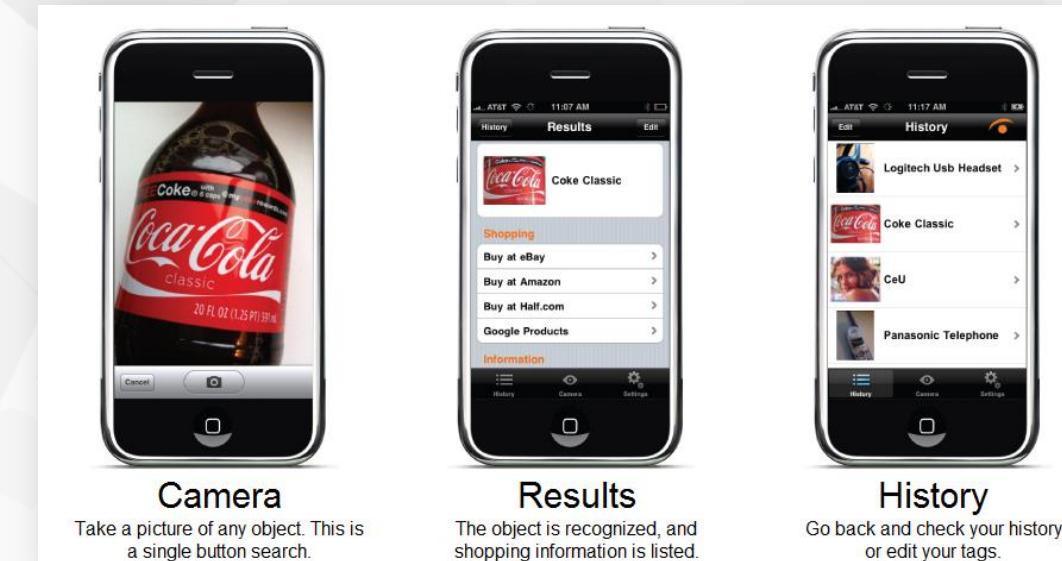
Plink

- Founded in April 2009 @ London with \$100K seed
- It stood out for its ability to quickly identify art with the snap of a photo
- Acquired by Google in April 2010



IQ Engine

- Founded in June 2008 @ Berkeley, CA
With \$3.8 million in Round 1
- It offers an image recognition platform called Glow that recognizes scenes, objects, landmarks, text and people in photos.
Used to automatically tags and organizes users' photos.
- Acquired by Yahoo! in Aug 2013;
Used to improve Flickr photo organization and search



Camera

Take a picture of any object. This is a single button search.

Results

The object is recognized, and shopping information is listed.

History

Go back and check your history or edit your tags.

LookFlow

- Founded in 2009 in Silicon Valley
- Build new ways for people to find, explore, collect, and share content they're interested in; using "an online search and discovery" technology
- Acquired by Yahoo! on Aug 2013;
To leverage its ML algorithms for Flickr



YAHOO!

LookFlow is joining the Flickr team at Yahoo!

We built LookFlow as an entirely new way to explore images you love — combining delightful user experiences with the latest advances in machine learning.

Flickr is the largest collection of images we love. They share our passion for creating phenomenal experiences & technology to help you discover those images.

We couldn't be more excited.

Fret not, LookFlow fans. Keep an eye out for our product in future versions of Flickr — with many more wonderful photos and all that Flickr awesomeness!

We'll also be helping Yahoo build a new deep learning group. If you're passionate about deep learning and want to help solve big problems, please [contact us](#).

Thanks to all the friends of LookFlow that have supported us. Special thanks to Michael Dearing, John Lilly, Reid Hoffman, Alex Rampell, Josh McFarland, Max Ventila, and Jeff Hammerbacher.

—Bobby, Simon and Team LookFlow

Kooaba

- Founded in Nov 2006 from ETH, Zurich;
\$2.9 million in 1 round;
- Image matching: mobile app that connects static printed newspaper format to dynamic online social sharing utilities
- Recently developed cloud-based image recognition solutions that integrate state-of-the-art visual recognition in applications
- Acquired by Qualcomm in Jan 2014;
used to boost its Vuforia augmented reality platform



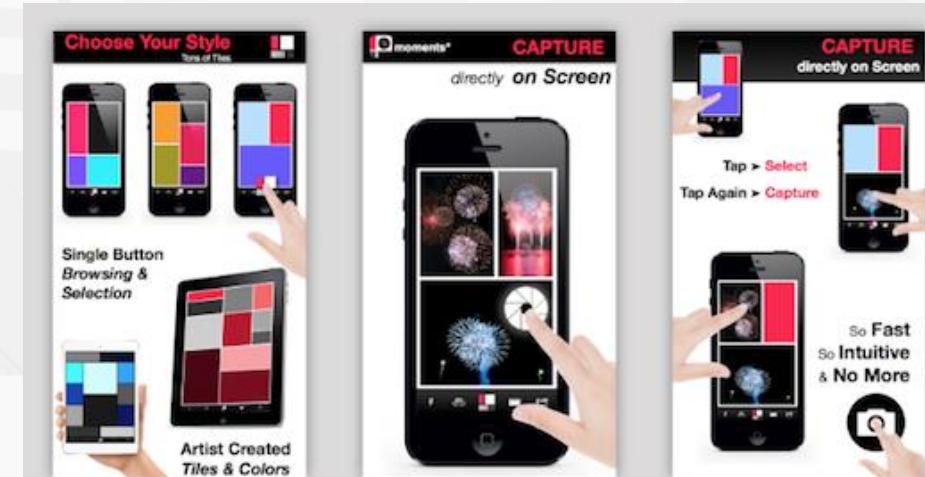
EuVision Technologies

- Founded in U of Amsterdam in 2010
- Technology: Image annotation
Launched mobile app named Impala that helps user organize their photos automatically into 20+ categories.
- Acquired by Qualcomm in Sept 2014
Doing innovative research to integrate image/video technologies into mobile space



Madbits

- Founded in New York
- It deploys deep learning technology to assign relevant information to raw images.
to automatically organize large databases of images.
- Acquired by Twitter in July 2014.
To help bolster its image search feature and catalogue the firehose of photos uploaded to the network



VisualGraph

- Visual analysis techniques to connect images into a visual graph – towards visual Web
- It deploys state-of-the-art machine vision tools, such as object recognition (e.g. shoes, faces), with large-scale distributed search and machine learning infrastructures

- Acquired by Pinterest in Jan 2014.

Recently announced products to offer recommendations to Pinterest users in fashion domain



Cloud-based, Big-data Image Recognition / Visual Search platform

Examples of commercial visual-search system
[Google Goggles, Search-by-image] [Amazon A9]

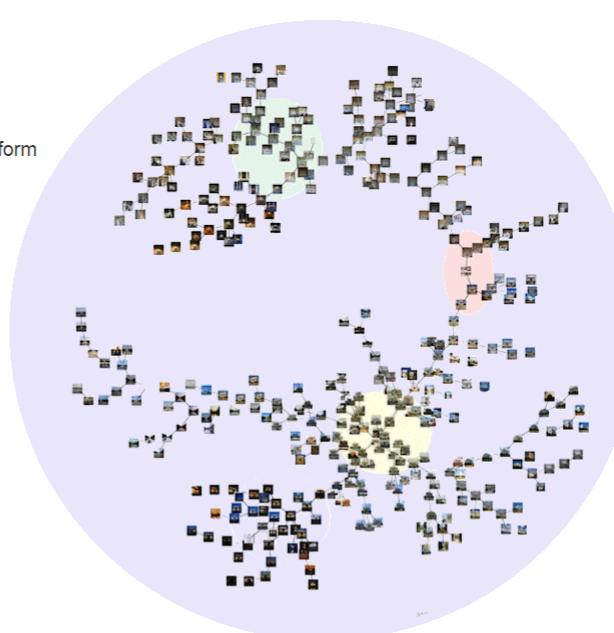
Our previous works on image graphs
[VisualRank For Product Images] [Google Image Swirl]

Technology
[Object Recognition]
- Face, People Detection (accuracy +98% of face.com)
- Cars, fashion objects, body, textured objects,

[Large-Scale Visual Search System]
- Product Recommendation
(Distributed, in-memory, visual-search system)

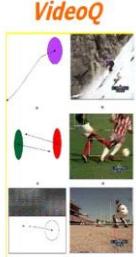
[Image Annotation]
- Deep learning, coming soon.

Team / Mission
Ex-Google machine vision scientist and engineers.
Connecting world's visual inspirations.



Deployment of Visual Technologies -1

Variety of Visual Technologies: Lots of Opportunities



Automatic Photo Tagging and Visual Image Search



Deployment of Visual Technologies -2

In General Commercial Visual Search Engines

SERVICE	TECHNOLOGY	TARGET USERS	BUSSINESS MODEL
Kooaba	Image recognition	Entertainment	SaaS-based
IQ Engines (oMoby)	Image recognition Crowdsourcing	Shopping	SaaS-based
Mobile Acuity	Image recognition	Consumers, Marketers	Contracting with marketers
LinkMe Mobile	Image and audio recognition	Consumers, Marketers	Contracting with marketers
Snaptell	Image recognition	Consumers, Marketers	Contracting with marketers
Point&Find (NOKIA)	Image recognition	Mobile users, Marketers	Engage more users in NOKIA experience
Gazopa	Image recognition	Shopping, Mobile users	Improving on-line shopping experience
Google Goggles	Image recognition	Mobile users	Advertising-based
Clic2c	Watermarking	Entertainment	Contracting with marketers
WeKnowIt IMG REC	Image recognition	Tourism	Touristic promotion actions
Wizup	Image and audio recognition	Consumers, Marketers	Contracting with marketers
TinEye Mobile (Snooth)	Image recognition	Wine industry	Advertising-based



Deployment of Visual Technologies -3

- Highlights of most current deployments (**mostly in B2B settings**):
 - Most products are based on image search, matching & annotation technologies
 - Some of the impressive systems are actually crowd-sourced-based
 - Visual recognition has just started
 - Hardly anyone with real-time in-video products
- Recent products are based on vertical domains, in four popular vertical domains:
 - Fashion
 - Home furnishing
 - Products
 - Surveillance

Deployment of Visual Technologies -4

Some vertical domain visual search engines

Company	Appln	Technologies
Snapfashion	Fashion find Apps	<ul style="list-style-type: none">• Similarity Search• App development
Superfish	Visual search API for image	<ul style="list-style-type: none">• Similarity Search
Cortica	Contextual Ads Serving for image	<ul style="list-style-type: none">• Matching-based recognition
Facec++	Face, then to finance & surveillance	<ul style="list-style-type: none">• Similarity search• Object recognition
Dextro	Visual recognition for live video streams	<ul style="list-style-type: none">• Visual analytics for images/videos
ViSenze	Visual Search & recognition API for image and video	<ul style="list-style-type: none">• Similarity search• Domain ontology• Object recognition



Highlights of ViSenze

ViSENZE

Simplifying the Visual Web

AWARDS



EMERGING
TECHNOLOGY



2014



2014



MOST INNOVATIVE
SOLUTION



Application Tools
& Platforms

VISUAL POWERED E-COMMERCE SITES

BUY SIMILAR FROM CLOZETTE SHOPPE

BUY SIMILAR is a fun tool that uses colours, shapes and patterns found in the photo to search for visually similar fashion items that you can buy at Clozette SHOPPE!

SEARCH SHOES



SEARCH DRESS



STEP 1: Drag the frame to determine which part of the photo you like to search for similar items.



STEP 2: Select a category to search

ALL DRESS TOP SKIRT PANTS BAG SHOES FIND NOW

O'SHARE

搜索时尚

Iris 結

追蹤

建議相似的單品

更多 >

立即上傳圖片，體驗圖片搜尋樂趣

樂天

搜索

立即上傳圖片

马上搜尋://

Home 頭條新聞

我傳上傳的圖片

AB-1 拍賣 - 480

AB-2 拍賣 - 480

AB-3 拍賣 - 480

AB-4 拍賣 - 480

AB-5 拍賣 - 480

NEXT

NEXT



**Snap
Search
Buy !**

*Shazam for
Fashion Commerce !*



Making Visual Content Targetable and Relevant to Advertisers

The screenshot shows a news article on the herworldplus.com website. The main headline is "She does it again: Duchess Kate's DVF wrap dress sells out online". Below the headline is a photo of Prince William and Duchess Kate at the Blue Mountains in Australia.

Standard Ad impression remains unchanged: A diamond ring advertisement for Lee Hwa Jewellery is displayed on the right side of the page. The ad features a large diamond ring and the text "Noble Diamond for Forevermark". A red bracket on the right indicates that the standard ad impression remains unchanged.

Contextual eCommerce Ad: A ZALORA advertisement for a blue and white patterned skirt is displayed below the main article. The ad includes the text "This week's specials". A red bracket on the right indicates that the fashion merchandise image is auto-selected or recommended based on closest similar outfit from a fashion database.

Key elements visible on the page include the herworldplus logo, navigation menu, and various news and shopping sections.



Contextual Advertising in Videos

Detect and recognize brands/logos in videos to associate with relevant Ads or Recommendations

In-Video Recognition



Use cases for Visual Search & Recognition

- Offline to Online
- Online to Offline
- Large-scale database retrieval
- Visual recommendation
- Supports multi-categories:
 - Fashion & related
 - Packaging
 - Home décor, furniture
 - Diagrams/drawings
 - And MANY more



Deployment of Current Visual Technologies

- Hardly anyone has deployable real-time in-video recognition technologies
- Video Recognition is still not widely used and talked about, even in research: *WHY?*

OUTLINE

- Retrospective on Image Search Since 2010
- Why Not Much Research in Video
- Some Current MM Research Efforts
- Summary

Current Needs of Industry

- Several urgent needs are all video based:
 - Visual recognition to in-video advertising
 - Live (video) event detection
 - Live Surveillance applications
- This can be seen in business/ industrial trends in social media..

Review of Social Media Platforms

- Social Network platforms
 - Three major platforms: Private, Professional, Public

facebook

LinkedIn

twitter

- Media Sharing platforms
 - YouTube, Instagram, Flickr, .. , Periscope, Vine



- Social Messaging Platforms
 - WhatsApp, LINE, Wechat, SnapChat, ...



★ Recent popular apps are all image/video based..

- Social Curation Platforms: Pinterest ...



Major Changes in Last few Years!

1. Image/Video handling

- Top recent social media platforms are all image/video centric



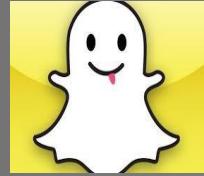
- Live video from GoPro, health sensors & appliances, etc

2. Live aspect is central now

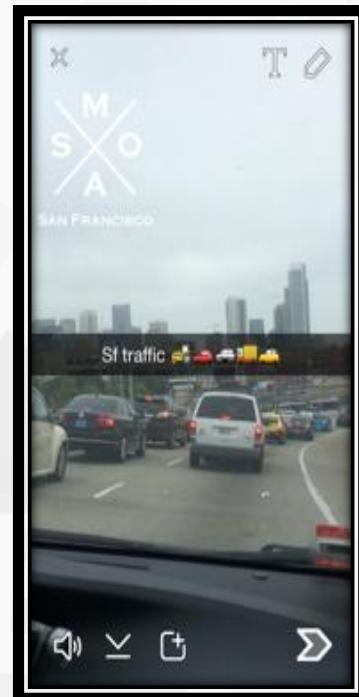
- With Live, comes the ability for continuous sharing and feedback..
- Users now want to get instant feedback: from friends and systems
- Problems in understanding and reconstructing contents in live contents, including videos



SnapChat -1



- A Multimedia Machine?
- ***Why it is so popular?***
- “Temporal Real Time Messages”: people decide how long others view their photos/videos
- Promotes creativity: Drawing, geo-filters & other ‘add ons’.

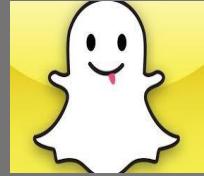


With one tap you can decide who sees the message.

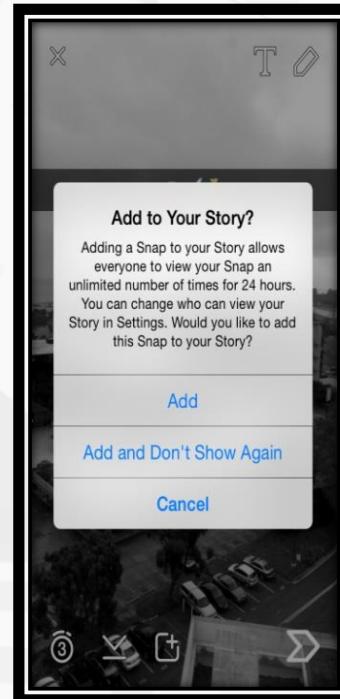


Depending on where you are, it provides geo-related filters; this adds another layer for communicating.

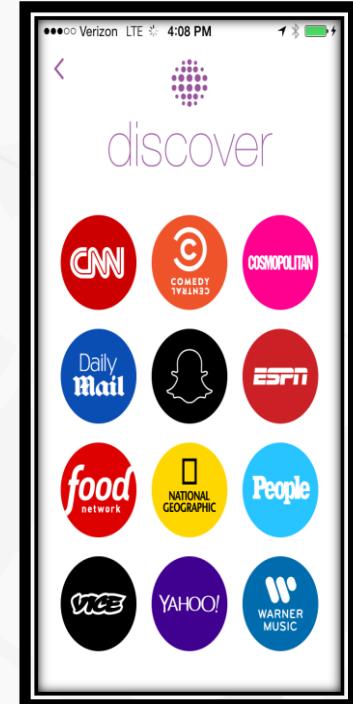
SnapChat -2



- ***Why it is so popular?***
- People can view stories from friends, contribute to nearby stories eg “NUS Graduation”, view live stories, e.g. Oscars, Sporting Events, and Fashion Shows.
- The discover feature allows people to view news from top sources, eg CNN, ESPN...



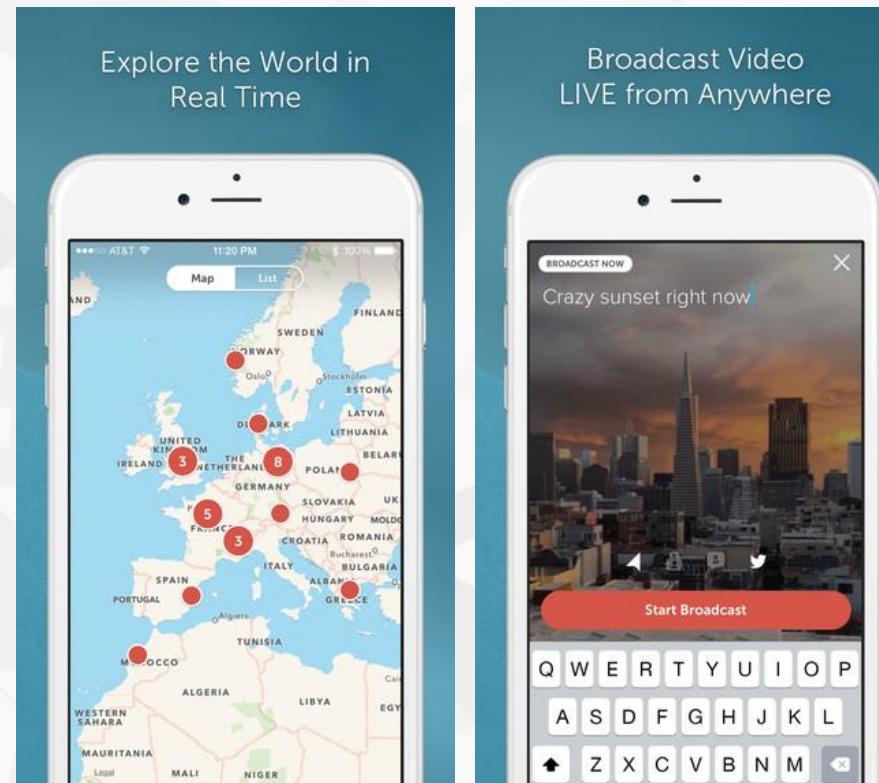
Creating a story, enables your friends to see what's happening in your life through photos and videos. **Paves the way for constant user engagement**



Users are able to consume content by brands. Opening the door for a way to consume news.

Periscope

- Similar to SnapChat in supporting live broadcast from mobile phones
- Users can broadcast and watch any other live broadcasts
- Why it is hot??
 - Support mobile-based broadcast
 - Track events live
 - Interactive
 - Integrate with Twitter Social Network



Very Little Research in Video -1

- Given the needs and popularity of video in industry, there are insufficient research efforts on video
- Conference statistics in MM-related conferences:

Conference	Total #	# that is visual-related	# related to Video
MM2014	59	46	15 (32%)
MM2013	53	38	13 (34%)
MM2012	67	50	14 (28%)
<hr/>			
ICMR2014	50	46	24 (52%)
ICMR2013	37	31	14 (45%)
ICMR2012	56	47	15 (32%)
<hr/>			
CVPR2014	540	378	113 (30%)
CVPR2013	471	267	91 (57%)
CVPR2012	465	275	79 (29%)



Very Little Research in Video -2

- Major Research topics (though many might not actually be in video)
 - Event Recognition/Detection
 - Action Recognition
 - Object Tracking
 - Saliency Prediction
 - Video Segmentation
 - Video Classification
 - Geographic Location Tagging
 - Video Copy Detection
 - Summarization

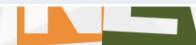
Very Little Research in Video -3

- What are the major obstacles?
 - **Lack of large-scale datasets?**
 - **Ability to process large-scale videos**
 - **Too obsessed with visual analysis?**

An overview of existing Video datasets

(complement from Yigang Jiang)

Dataset	# Videos	# Classes	Year	Manually Labeled ?
Kodak	1,358	25	2007	✓
MCG-WEBV	234,414	15	2009	✓
CCV	9,317	20	2011	✓
UCF-101	13,320	101	2012	✓
THUMOS-2014	18,394	101	2014	✓
MED-2014	≈28,000	20	2014	✓
Sports-1M	1M	487	2014	✗
.....				



Recent Datasets

Dataset	Size	Source	Labels	Def. of Label	Publicly Accessible
MED	~0.3 M	Youtube	40 Events	Self-Defined	~18% (released 2014)
FCVID	91,223	Youtube	239 Categories	Self-Defined	100% (released 2015)
NUS/ CMU/ Yahoo!	106,000 (to extend to 0.7 M)	Flickr (consumer videos)	420 Concepts + 100 Scenes	ImageNet + SUN	100% (to be released soon)

FCVID: Fudan-Columbia Video Dataset

- One of the **largest** public benchmark Internet videos with manual annotations (**239** categories)
- Covering many categories organized in a **hierarchical** structure
- Released in Feb. 2015!
<http://bigvid.fudan.edu.cn/FCVID/>

Fudan-Columbia Video Dataset (FCVID)

(239 categories; 11 higher level groups; 32 second level groups)

All categories (254)										
Sports (46)						Music (17)			DIY (21)	
Sports Amateur (14)	Sports Professional (20)		Extreme Sports (8)	Sports for the disabled (4)		musical performance without instruments (6)	solo musical performance with instruments (10)	group musical performance with instruments (3)		
baseball	baseball	marathon	rock climbing	wheelchair basketball	singing on stage	guitar performance	symphony orchestra performance	making rings	Making wallet	
basketball	basketball	Rhythmic gymnastics	skateboarding	wheelchair tennis	singing in ktv	piano performance	rock band performance	making earrings	Making pencil cases	
soccer	soccer	taekwondo	surfing	wheelchair race	beatbox	violin performance	chamber music	making bracelets	Making phone cases	
biking	biking	archery	skydiving	wheelchair soccer	chorus	accordion performance		making festival cards	Making photo frame	
ice skating	swimming	fencing	bungee jumping			cello performance		making clothes(sewing)	Making bookmark	
swimming	skiing	Car racing	rafting			flute performance		making a paper plane	Building a dog house	
skiing	American football	rowing	parkour			trumpet performance		making paper flowers	Blowing up an air bed	
American football	tennis	sumo wrestling	kitesurfing			saxophone performance		knitting	Pitching a tent	
tennis	sports track	diving				harmonica performance		assembling a computer	Changing tires	
table tennis	boxing	shooting				drumming		assembling a bike	Making shorts	
badminton									Tying a tie	
billiard										
frisbee										
shooting										



Fudan-Columbia Video Dataset (FCVID)

(239 categories; 11 higher level groups; 32 second level groups)

All categories (254)									
Beauty & fashion (10)		Cooking & Health (30)			Leisure & Tricks (22)			Art (10)	
Beauty(6)	fashion(4)	Food (13)	Drinks (5)	Health(12)	Leisure sports(7)	Common leisure activities(11)	Tricks(4)		
making up	showing fashionable high heeled shoes	barbecue	making coffee	Dumbbell workout	roller skating	flying kites	yoyo tricks	painting	
make lipstick	showing fashionable handbags	making French fries	making tea	Barbell workout	fishing	bumper cars	pen spinning	sculpting	
eye makeup	fashion show	making sandwich	making juice	punching bag workout	boating	kicking shuttlecock	solving magic cube	doing graffiti	
hair style design	red carpet fashion	roasting turkey	making milk tea	push ups	golfing	playing chess	card manipulation	making ceramic craft	
nail art design		making sushi	making mixed drinks	pull ups	bowling	playing bridge		solo dance	
face massage		making salad		sit ups	hiking	snowball fight		group dance	
tattooing		making pizza		rope skipping	horse riding	making a snowman		social dance	
		making cake		treadmill		arm wrestling		spray painting	
		making hotdog		Hula hoop		playing with Nun Chucks		sand painting	
		making cookies		jogging		playing with remote controlled aircraft		Chinese paper cutting	
		making ice cream		yoga		playing with remote controlled cars			
		making Chinese dumplings		Tai Chi Chuan					
		making egg tarts							



Fudan-Columbia Video Dataset (FCVID)

(239 categories; 11 higher level groups; 32 second level groups)

All categories (254)															
Everyday Life (54)								Nature (26)				Travel (11)		Tech & Education (7)	
Places (6)	Activities(4)	Kids (7)	Family Events (4)	Chores(8)	Social Events (10)	Public Events (7)	Pets and others (8)	Sceneries (8)	Natural Phenomenon(7)	Animal (11)	Transportations(4)	Tourist Spots(7)	High-tech product introductions (6)	Educatio n (1)	
Temple exterior	hair cutting	Kid playing on playground	birthday	cleaning windows	wedding ceremony	parade	bird	beach	sunset	dolphin	train	Egyptian pyramids	psp	classroo m	
bridge	shaving beard	kids building blocks	family dinner	cleaning floor	wedding reception	fire fighting	dog	mountain	tornado	turtle	airplane	Eiffel tower	smart phone		
Cathedral exterior	brushing teeth	kindergarten	decorating Christmas tree	weeding	wedding dance	car accidents	cat	river	lightning	snake	ship	the great wall	panel computer		
museum interior	walking with a dog	baby eating snack	camping	washing dishes	graduation	street fighting	hamster	waterfall	sandstorm	spider	bus	the Statue of Liberty	single-lens reflex camera		
library interior		baby crawling		car washing	dinner at restauran	public speech	rabbit	forest	volcano eruption	cow		the oriental pearl TV tower	telescope		
amusement park		washing an infant		fruit tree pruning	picnic	fireworks show	delicious food	ocean	solar eclipse	panda		the Leaning Tower of Pisa	laptop		
		kids making faces		shoveling snow	group banquet	car exhibition	house plants	desert	lunar eclipse	butterfly		Taj Mahal			
				cleaning carpet	debate		toy figures	grass land				camel			
					drinking inside bar							gorilla			
					dancing inside nightclub							giraffe			
												elephant			



FCVID: Example 1



Everyday life

Kids

Kids playing
blocks



FCVID: Example 2



Sports
↓
Sports
Professional
↓
Boxing



NUS-CMU-Yahoo! Dataset

- Targets on consumer videos, mostly recorded by mobile devices
- Source and Scale: Yahoo Creative Commons 100M
 - 0.7 million videos from Flickr
- Targeted Annotations
 - 420 entry-level concepts from ImageNet, 100 scenes from SUN
 - To release tagging for 0.1M videos soon, the rest later



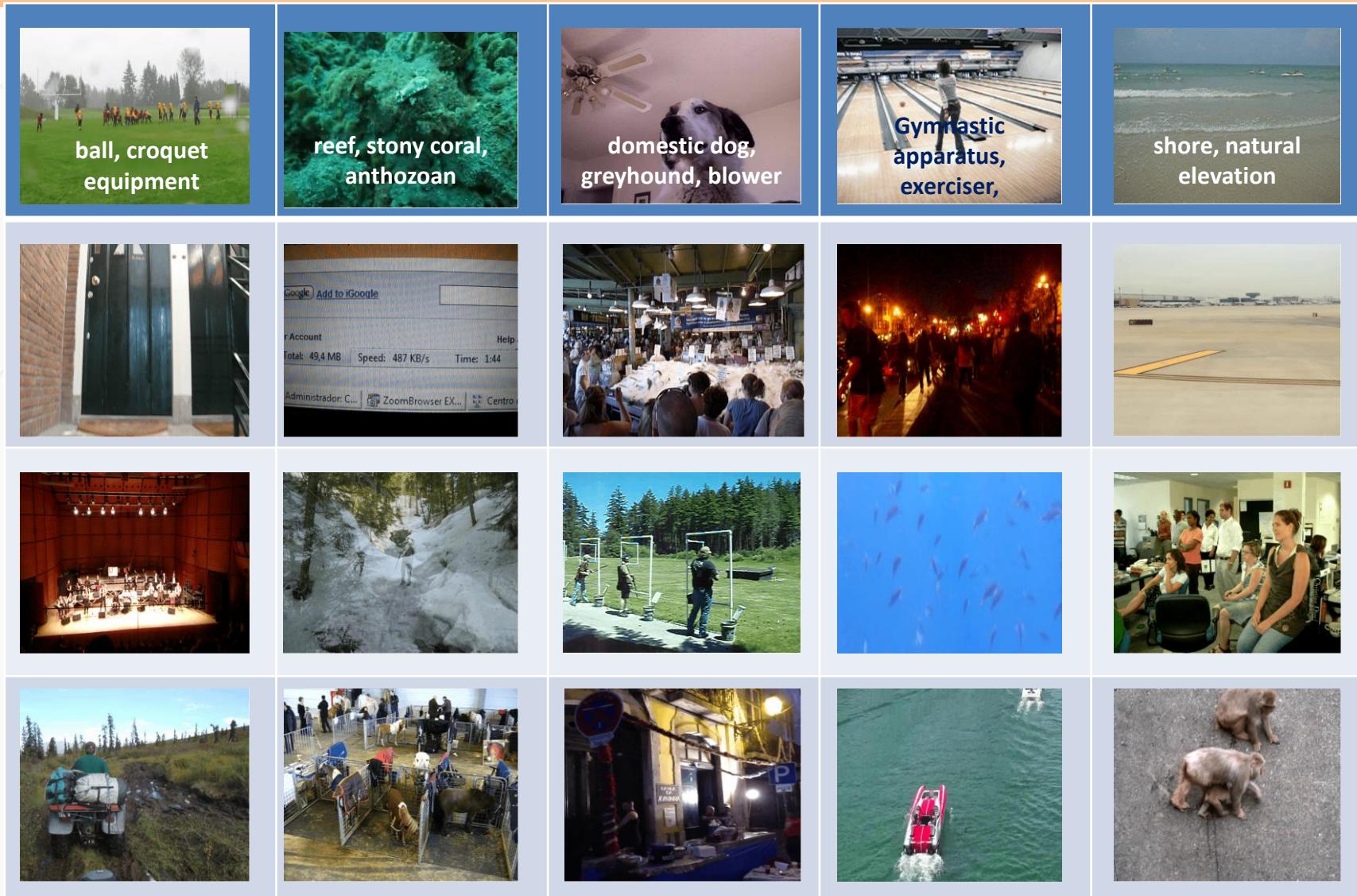
Concept Definitions

dog	flower	ball	wheel vehicle	vessel, water-craft	fish	insect	kitchen appliance	Computer	gun
boxer, bull mastiff, great dane, husky, saint bernard, collie, kelpie, shetland, toy terrier, japanese spaniel,c ihuahu a, gordon setter, english setter	daisy, yellow lady's Slipper	ping- pong ball, volleyb all, rugby ball, punchin g ball, soccer ball, tennis ball	Tricycle, motor scooter, unicycle	trimaran, catamaran, gondola, fireboat, submarin e, canoe, yawls, speedbo at, lifeboat, liner	coho, lionfish, garfish, barracuda, pufferfish, blowfish, eel, rock beauty, tench, goldfish, anemone fish	cricket, grasshopper, bee, ant, mantis, cockroach, walking stick, lacewing, fly, leafhopper, cicada, dung beetle, rhinoceros beetle, carabid beetle	microwa ve, toaster, waffle iron	web site, notebook, laptop, hand- held computer, desktop computer	six- shooter, assault rifle, cannon, rifle

Scene Definitions

INDOOR:				OUTDOOR NATURAL:			OUTDOOR MAN-MADE:		
shopping and dining	workplace	home or hotel	transportation	water, ice, snow	mountains, hills, desert, sky	forest, field, jungle	sports fields, parks, leisure spaces	houses, cabins, gardens, and farms	commercial buildings
banquet_hall, beauty_salon, Bookstore, anechoic_chamber, butchers_shop control_room , op, candy_store Corridor, dentists_office , engine_room, coffee_shop e, , , engine_room, jewelry_shop hospital_room , m, music_store, veterinarians , _office Pharmacy, barrel_storage , ge	Attic, Bedroom, home_office, living_room, Nursery, Pantry, utility_room, barrel_storage	bus_interior, engine_room, limousine_interior, subway_station /platform, van_interior,	Natural, Dock, ice_floe, Iceberg, Pond, Sandbar, ski_slope, Swamp, underwater/coral_reef	Badlands, desert/vegetation	field/cultivate d, field/wild, golf_course, Pasture, putting_green , Rainforest, rice_paddy, Swamp	golf_course, hot_tub/outdoor, labyrinth/outdoor, Pavilion, putting_green, Racecourse, rope_bridge, ski_slope, stadium/baseball,	balcony/exter ior, doorway/out door, field/cultivate d, greenhouse/o utdoor, House, hunting_lodge/outdoor, outhouse/out door, Patio, rice_paddy	Alley, apartment_b uilding/outdo or, balcony/exter ior, Crosswalk, doorway/out door, market/outdo or, office_buildin g, phone_booth ,	Skyscraper

NUS-CMU-Yahoo! Dataset: Examples



NUS-CMU-Yahoo! Dataset: cont.

- Source and Scale: Yahoo Creative Commons 100M
 - 0.7 million videos from Flickr
 - To tag 420 entry-level concepts from ImageNet, & 100 scenes from SUN
 - To release tagging for 0.1M videos soon, the rest later
 - Support annotation, retrieval & video description tasks
- Plan to launch in MMM and/or ACM ICMR:
 - For interactive tasks and grand challenge tasks
- Encourage collaborations:
 - To tag other entities, and define/lead other tasks, like event detection

NOTE:

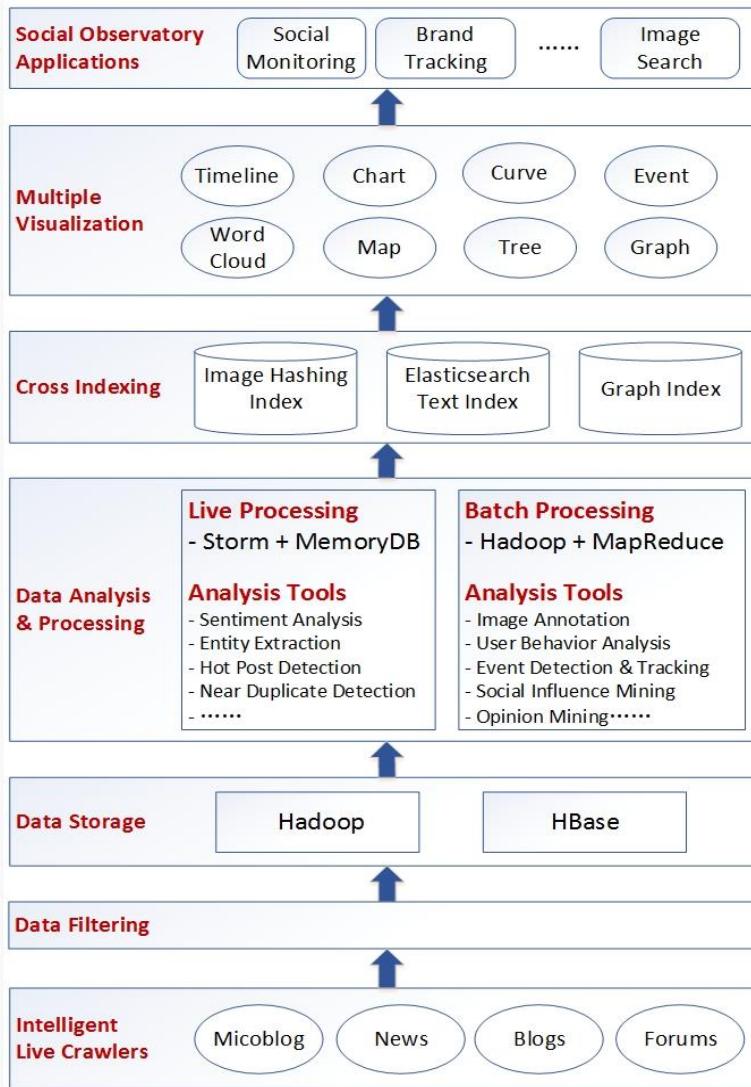
1. **Size:** largest when fully tagged
2. **Consumer-level:** first 100% consumer-level dataset
3. **Definition of Labels:** based on current popular ontologies;
can directly use off-the-shelf concept or scene detectors
pre-trained on ImageNet and SUN.



2. Ability to Process Large-Scale Videos

- Availability of Dataset is one of the main academic problem, but not the only problem
- The ability to process large-scale video is another:
 - Infrastructures for large-scale stream data
 - Technologies to process video in real-time
 - Mostly engineering tasks done in industry
- Works in such robust academic systems are beginning
 - For example, work by CMU in this conference: “Content-based video search over 1 million videos with 1 core in 1 second”
 - Need industrial scale infrastructures and expertise
 - Need technologies for visual audio and text analysis

NExT Architecture for Big Data Processing



- Back-End Tech Highlights
 - Fully automated
 - Cloud-based
 - 80 VMs
 - Big data
 - >100 TB, 2.6 billion records
 - Live + Batch Processing
 - Storm + Hadoop + Hbase
 - Distributed
 - Crawlers, Database, File Storage, Processing
 - Cross indexing
 - Elasticsearch index for text
 - Own hash-based image index
 - Graph index



3. Needs Multimedia Approach to Solve Real Problems

- All B2C apps use simple but reliable contextual info to drive applications
 - Mostly are ideas to help users communicate and have fun
 - For example, Geo-filter, Live Story and Discover features in SnapChat rely on manual curation and geo-tags
 - User profiling in social media products relies on check-in POI, user-defined profile, messages and images shared etc.
 - **What visual techniques are needed?**
- In contrast, B2B apps require to offer solutions to customers' problems
 - Most need technologies as basis/ barrier to entry



3. Needs Multimedia Approach to Solve Real Problems

- Because of the above, not many B2C ideas would come from academic community
- Our evaluation system values validated ideas with thorough evaluations, which could be the obstacles
 - This is especially the case when acceptance ratio is low
 - Hence hard for novel ideas to flourish..
 - We need to be conservative in pursing ideas – in name of publications
- How to address this problem?
 - Well recognized in ACM
 - We tried various means, including the introduction of “Brave New Idea” track



OUTLINE

- Retrospective on Image Search Since 2010
- Why Not Much Research in Video
- Some Current MM Research Efforts
- Summary

Need to Tackle Real Multimedia Problems

- Many real practical problems require multimedia, not visual, solutions
 - We need to make judicious use of text, image/video, location data etc.
- As illustration, I look at solutions to several problems using multimodal data:
 - MED Event Detection
 - Social TV
 - User Mobility
 - Event-Shop (Ramesh Jain)
 - Wellness

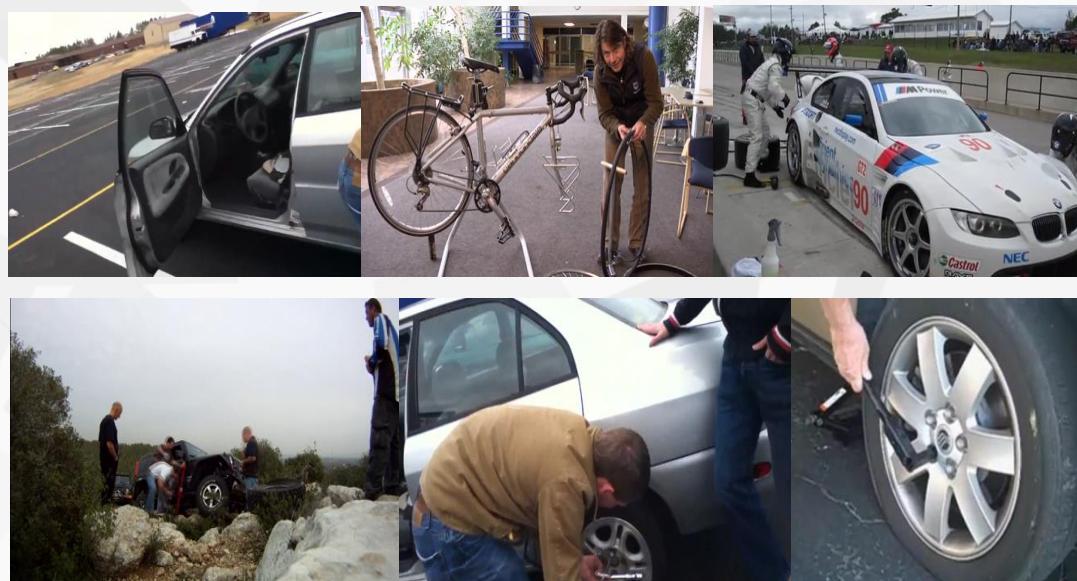
1) Event Detection in Video

- Given a video, determine if an event has occurred
- One key problem is to identify what the event-related concepts
 - Many current solutions employ complex visual analysis and corpus correlations to determine important related concepts
- For example,

event – Changing a vehicle tire

semantic concepts -

vehicle,
bicycle wheel,
car wheel ...



Why not use domain knowledge: From FrameNet Project

- What is Frame?

A *frame* is a data-structure for representing a stereotyped situation, like going to a child's birthday party

- **top levels** of frame:
situation (e.g. going to a child's
birthday party)
- **lower levels** of frame:
specific instances or data (e.g. *song*,
birthday song, *games*, *cake*, *party-meal* ...)

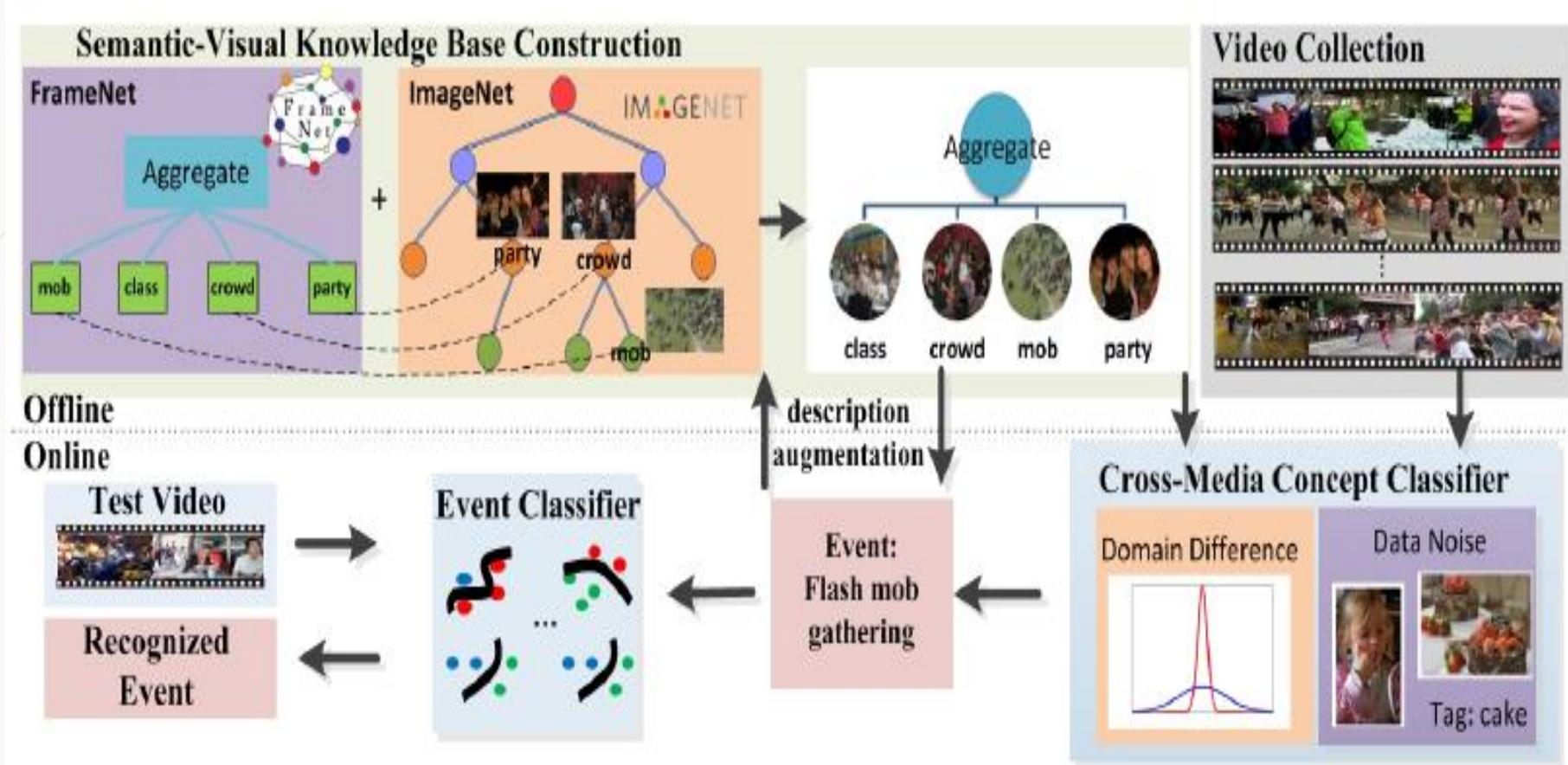
- Widely used in NLP research

[Marvin Minsky, A Framework for Representing Knowledge, 1974]

Frame	FrameNet LUs
Social_Event	banquet.n dance.n dinner.n party.n celebration.n ...
Forming_Relationships	betrothal.n divorce.n engagement.n wedding.n ...
Personal_Relationship	romance.n spouse.n marriage.n husband.n engagement.n...
Vehicle	bicycle.n bus.n cab.n car.n convertible.n limousine.n ...
Vehicle_Subpart	engine.n seatbelt.n tire.n wheel.n windshield.n ...
Leadership	premier.n prince.n president.n lawmaker.n...
Grooming	comb.v shampoo.v wash.v shower.v soap.v...

System Overview

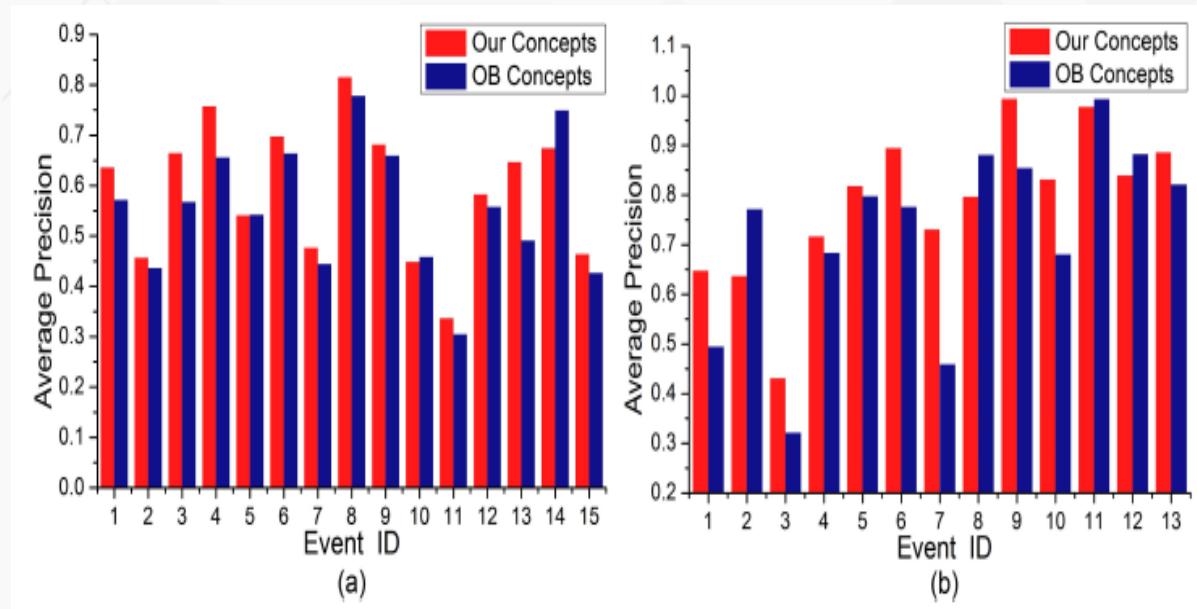
- Event: Flash mob gathering



Effects of Event-related Concept Set

- Video Event Recognition

Comparison with OB concepts [1] consisting of 177 most frequent objects on: (a) MED11, and (b) EVVE



- [1] L.-J. Li, H. Su, L. Fei-Fei, and E. P. Xing. Object bank. NIPS, 2010.
[2] J. Revaud, and H. Jegou. Event retrieval in large video collections with circulant temporal encoding. CVPR, 2013.

To appear in
IEEE ToM

2) Social TV

- Utilize second screen to offer simultaneous social media discussions on live broadcast TV shows:
 - to provide social and interactive features simultaneously with the TV viewing experience.



Live Social Media Analytics

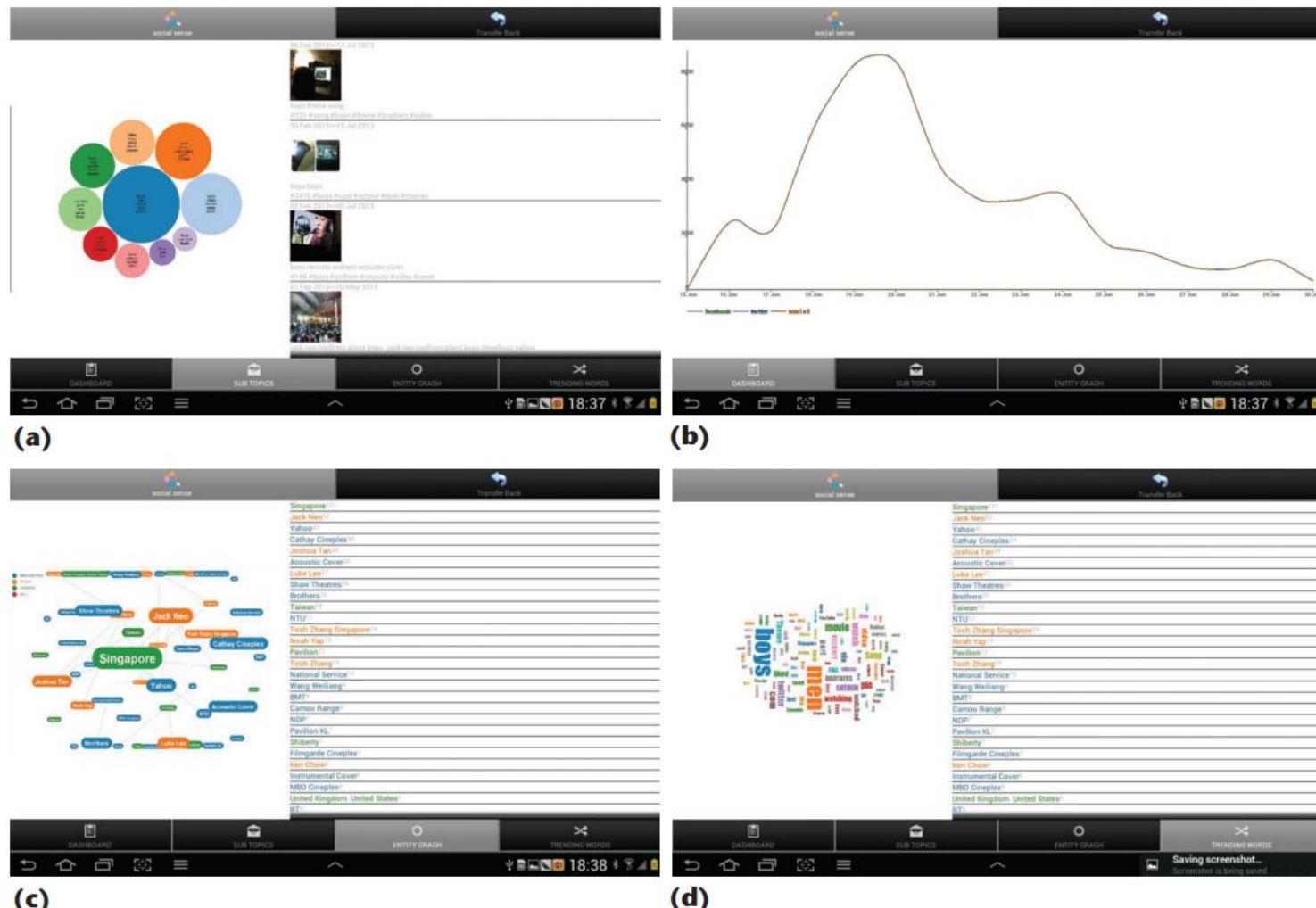


Figure 5. Social sense features. (a) Topics, (b) degree of interest, (c) entity graph, and (d) keyword cloud.

2) Social TV

- Utilize second screen to offer simultaneous social media discussions on live broadcast TV shows:
 - to provide social and interactive features simultaneously with the TV viewing experience.

Selected as Best Paper for
IEEE MM 2015

H Hu, et al. (2014): Toward Multiscreen Social TV with Geolocation-Aware Social Sense. IEEE MultiMedia 21(3): 10-19

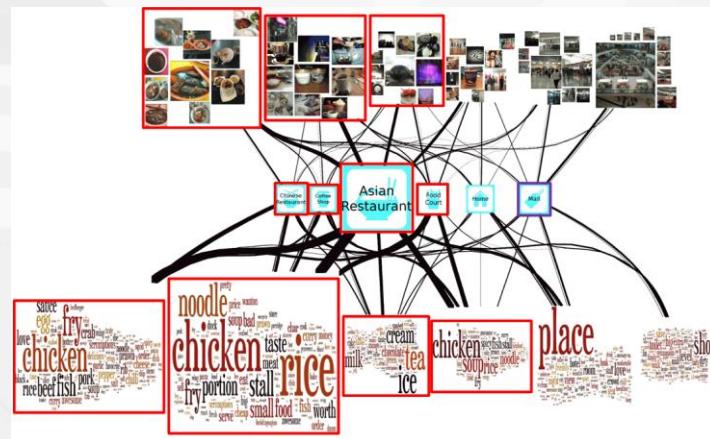


3. User Mobility

- Utilize data from multiple social media sources
- Mine relations between check-in venues and UGCs
- Identify landmarks of local venues
- Identify popular trials (of individuals and their friends)
- Analyze user demographic and interests communities
- Generate multi-faceted user relation graph



Travel paths and flows of users

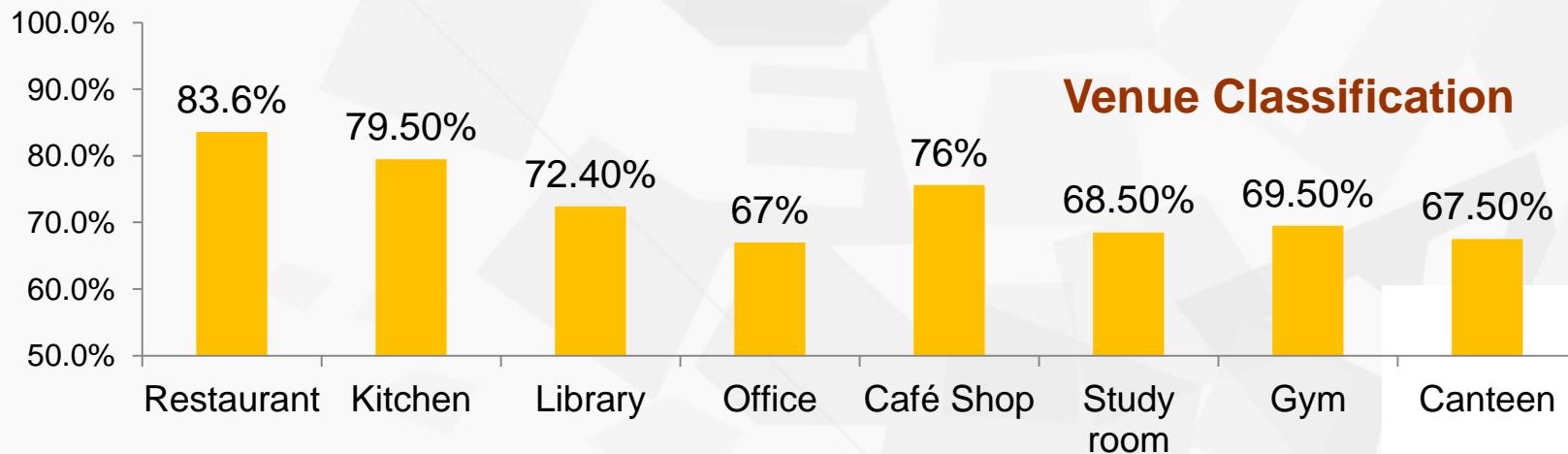


Interest community of food lovers in Singapore

Multimodal Location Estimation

- **POI Estimation**
 - GPS trace of users, and messages if any
 - Estimate the POI's that users have visited

- **Mapping to POI (Point of Interests)**
 - Use nearby venues found in 4Square, and other info
 - Utilized audio signals to estimate venue categories
 - Essential for micro-videos in social media sites



- **Classifier: POI of messages**

User profile: Mobility + Demography

- **Estimate User Demography**

- Utilize multi-source social media data, & if available, mobile phone traces
- For Age and Gender – to over 80% accuracy

User profile

Mobility profile

Demographic profile

Location
preference

Movement
patterns

Age

Gender

Personality

Occupation

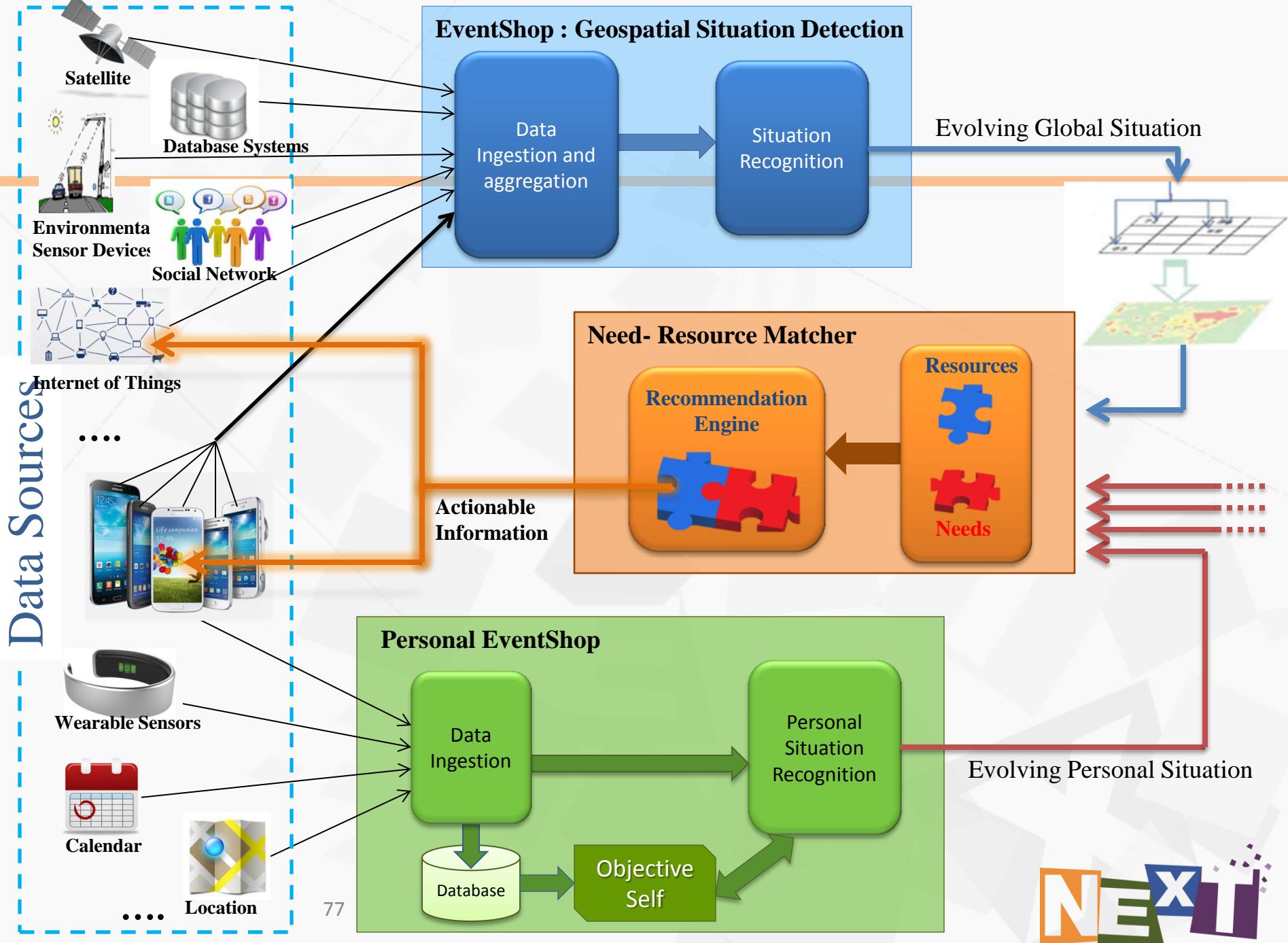
- To present in this conference on Thu afternoon

4) Geospatial Interpolation Analytics

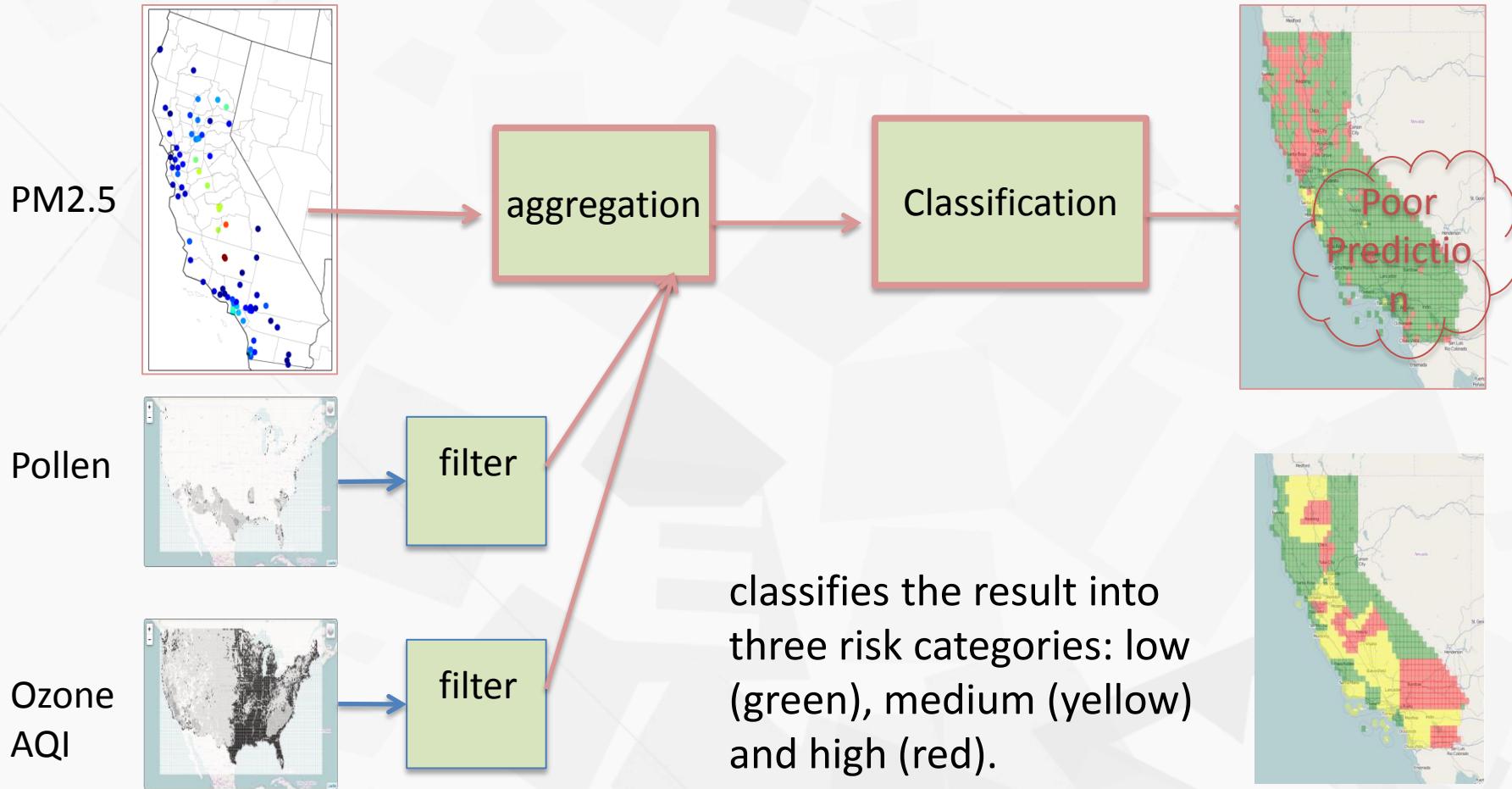
“Social Life Network”

Providing right RESOURCES
at the right TIME
in the right PLACE





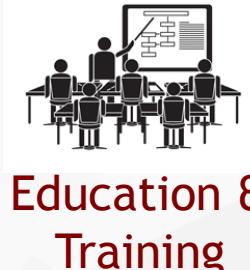
Asthma Risk Management Application



Geospatial Interpolation Analytics for Data Streams in EventShop
Nominated as Best Paper Candidate in ICME 2015

5) Computational Wellness

healthSenze gathers & organizes big-data from multiple sources



- Personalized wellness with big data
- Focus on Critical illnesses
- Predictive and prescriptive wellness



OUTLINE

- Retrospective on Image Search Since 2010
- Why Not Much Research in Video
- Some Current MM Research Efforts
- Summary

Summary

- Current states of MM applications:
 - Technologies in media search have evolved and matured to tackle applications in certain vertical domains
 - Wide range of data, knowledge sources and social network data available online – for various wide range of applications
- Why little research in video?
 - Live online videos feature in most recent popular social apps
 - Lack of datasets and infrastructures?
 - Need to look into multi-modal approach to tackle problems
- Why lack of good industry ideas:
 - Technical research supports good B2B app
 - But not B2C which based on ideas leveraging robust technologies
 - Academic research not gears towards that
 - Need fundamental change of how to promote novel ideas



Summary

- To address problems raised:
 - Work on truly multimedia data, not just visual
 - Work on some risky new ideas, not just matured formula
 - Work on large-scale real-life (end-to-end) problem

THANKS

Visit our Web Observatory:
<http://WWW.NEXTCENTER.ORG/>



The screenshot displays the homepage of the Live Observatory. At the top left is the logo "Live Observatory". At the top right are links for "Home" and "About". Below the header is a horizontal red line. Underneath the line are five circular icons, each representing a different sense module: "Live Crawler" (a globe with a magnifying glass), "Location Sense" (a globe with a location pin), "People Sense" (a person under a yellow beam of light from a blue UFO-like shape), "Topics Sense" (two people talking with speech bubbles), and "ORG Sense" (a network of nodes). Arrows point from the names below each icon to the respective icons. Below these icons is a large blue rectangular box. Inside the box, the word "ORG Sense" is at the top, followed by a horizontal line. To the left of the line is a red circle containing the white text "Try it!". To the right of the line is a paragraph of text: "ORGSense takes care of the WWW of your organization, in a word, what people are saying about your organization, who these people are, and where they are."

ACM ICMR Conferences

- ICMR2016: NYC, USA
 - John Smith (IBM Research), John Kender (Columbia Univ.)
- ICMR2017: Bucharest, Romania
 - Nicu Sebe (Univ. Trento), Bogdan Ionescu (Univ. Politehnica Bucharest)
- ICMR 2018 will be in Asia:
 - Call for organization out soon, please prepare your bid to support the conference

