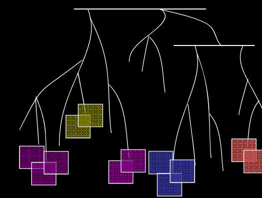
+



Exploring Image, Video and Multimedia in Large Data Applications

Prof. Shih-Fu Chang

http://www.ee.columbia.edu/dvmm

January 29th, 2016

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

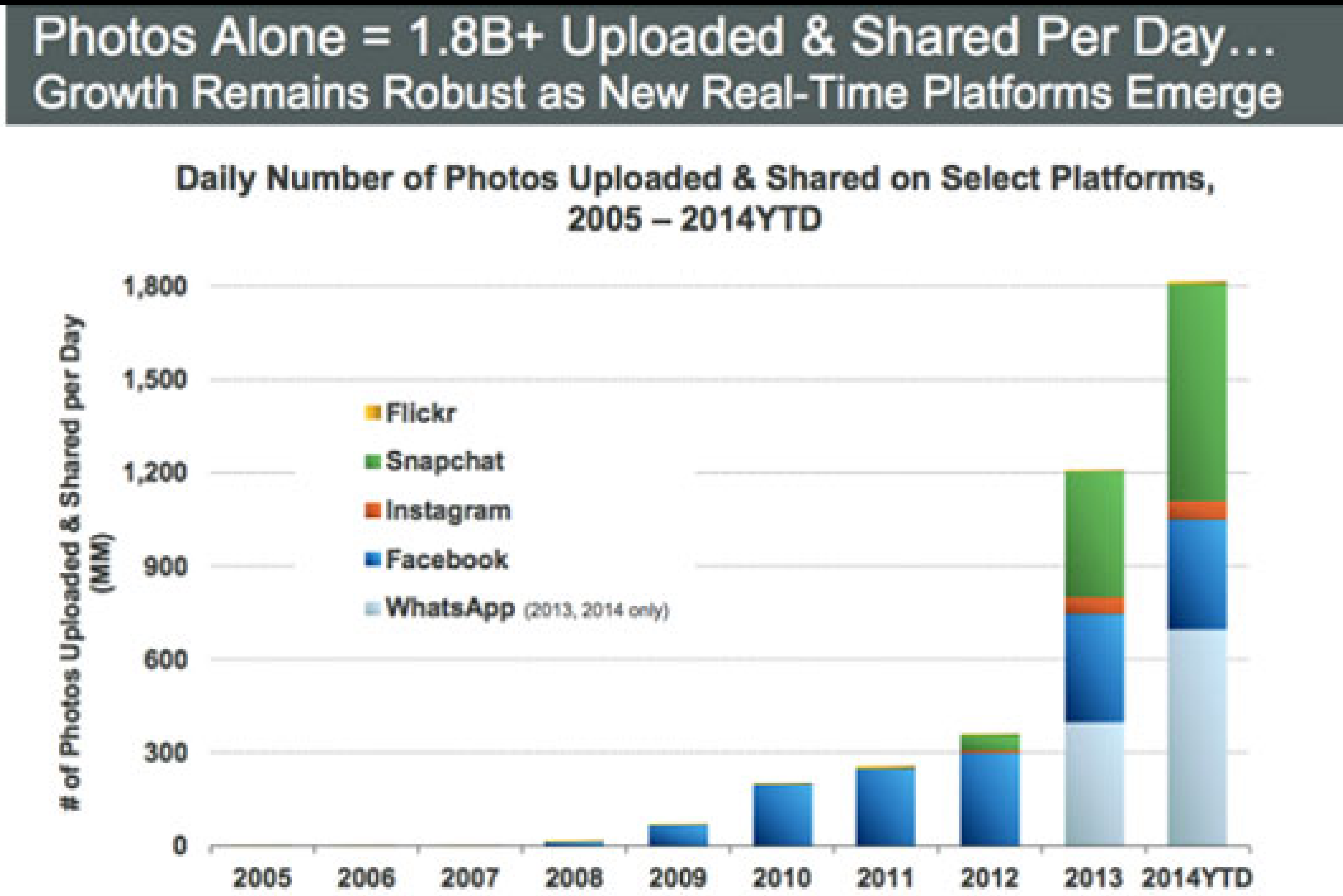
First Digital Camera in 1975
- film-less photography

by Steve Sassan of Kodak

- CCD array, A/D converter, 16 batteries
- 23 seconds to record a photo to cassette
- customized reader on a B/W TV for viewing

Q: quality, size, cost, store, share?

Today …



Source: KPCB estimates based on publicly disclosed company data, 2014 YTD data per latest as of 5/14

The New York Times

Researchers Announce Advance in Image-Recognition Software

By JOHN MARKOFF NOV. 17, 2014

FORTUNE

How Facebook is teaching computers to see

by Stacey Higginbotham @gigastacey JUNE 15, 2015, 1:15 PM EST

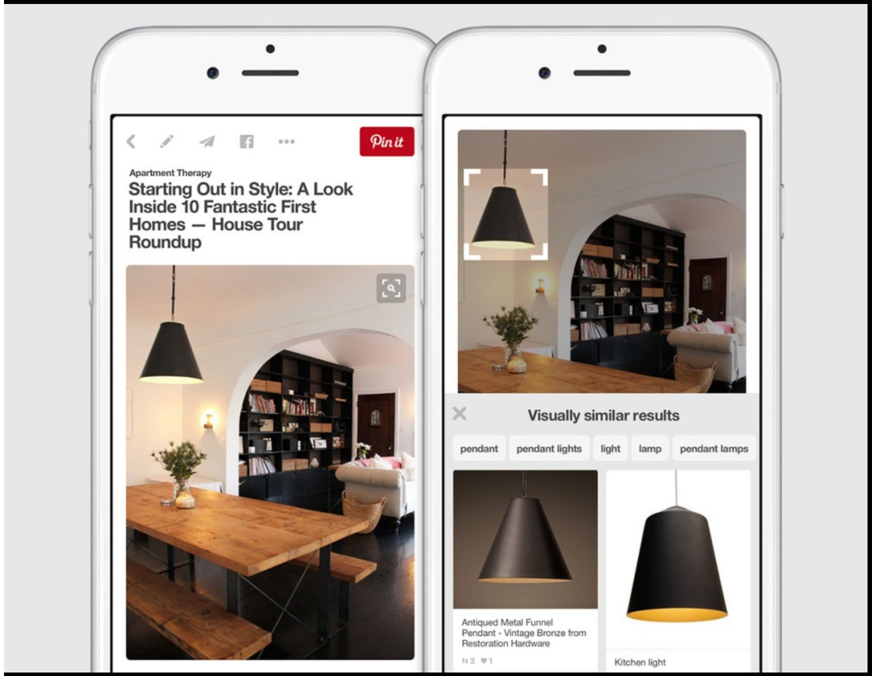
MIT Technology Review

AI Advances Make It Possible to Search, Shop with Images

Deep learning software has dramatically improved image recognition tools. Pinterest and Shoes.com are testing it out on shoppers.

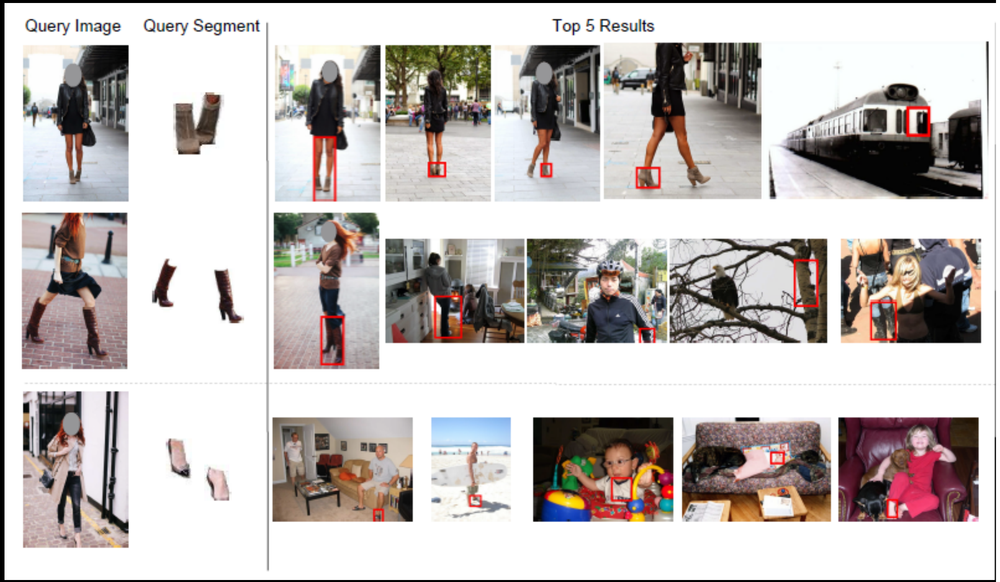
By Tom Simonite on November 17, 2015

Pinterest Visual Search



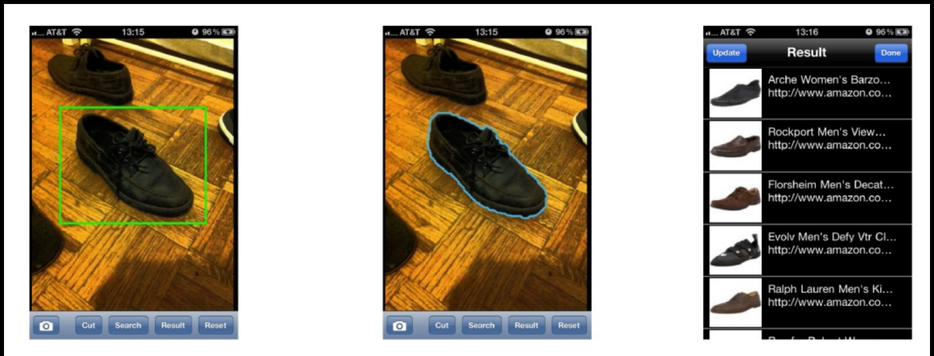
Instance Search

(R. Tao, A. Smeulder and S.F. Chang, CVPR 15)



Columbia Mobile Visual Search

(MM'11, 1KB compact hash code over 0.5M products)



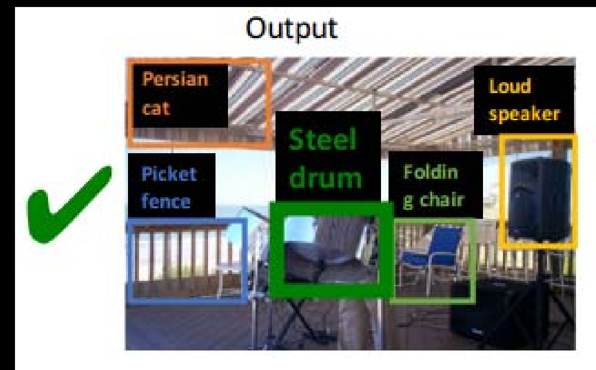
video demo

ImageNet:
Recognize > 1,000 object categories

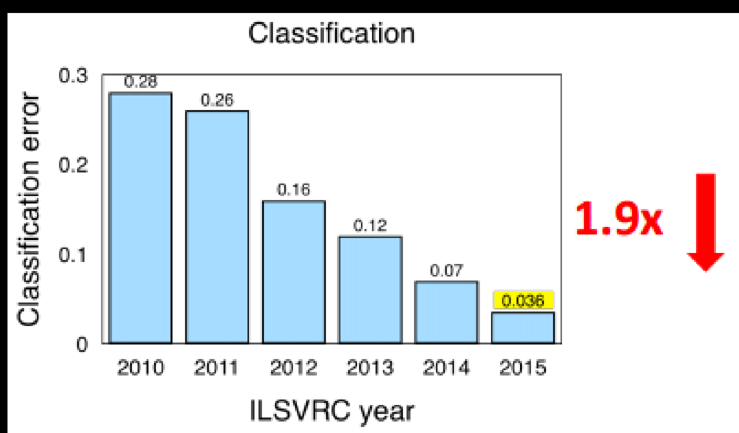
S.-F. Chang

Rapid Progress in Recent Years:

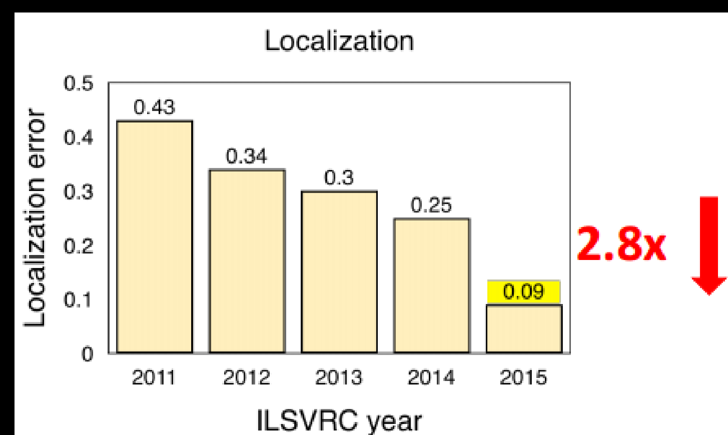
Classification/ Detection Task



$$\text{Error} = \frac{1}{100{,}000} \sum 1[\text{incorrect on image i}]$$



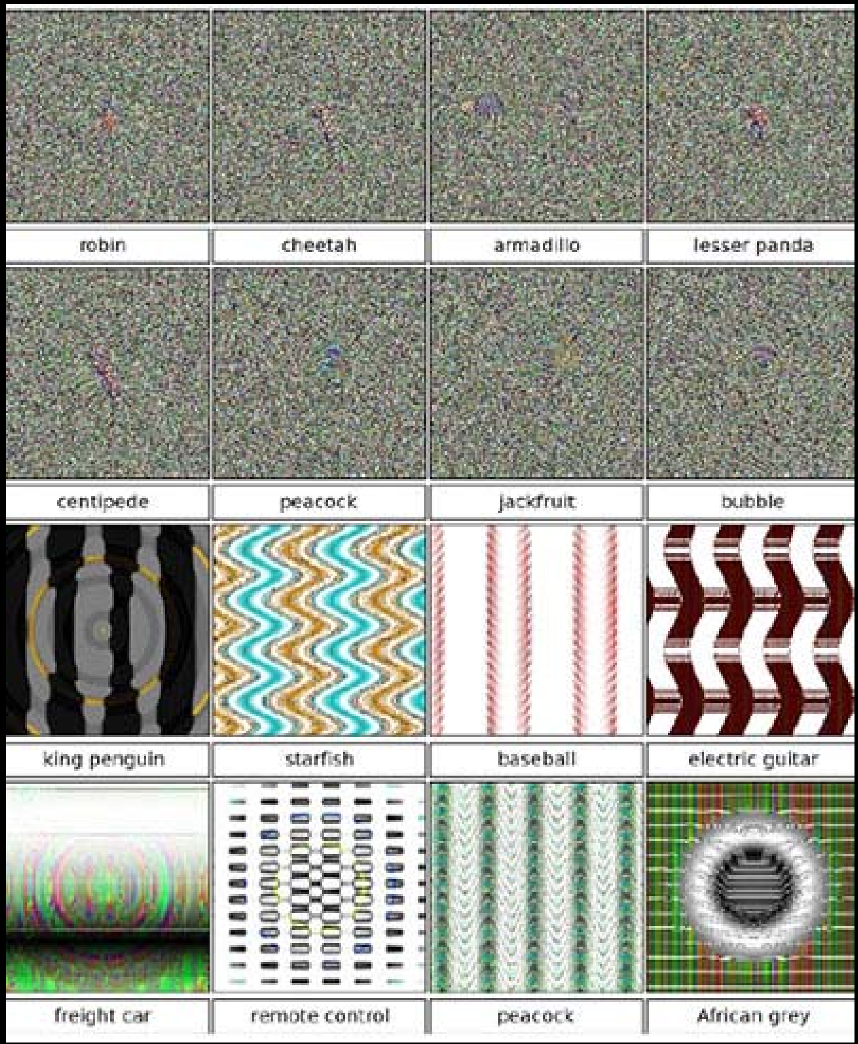
1.9x



2.8x

But Use with Caution!

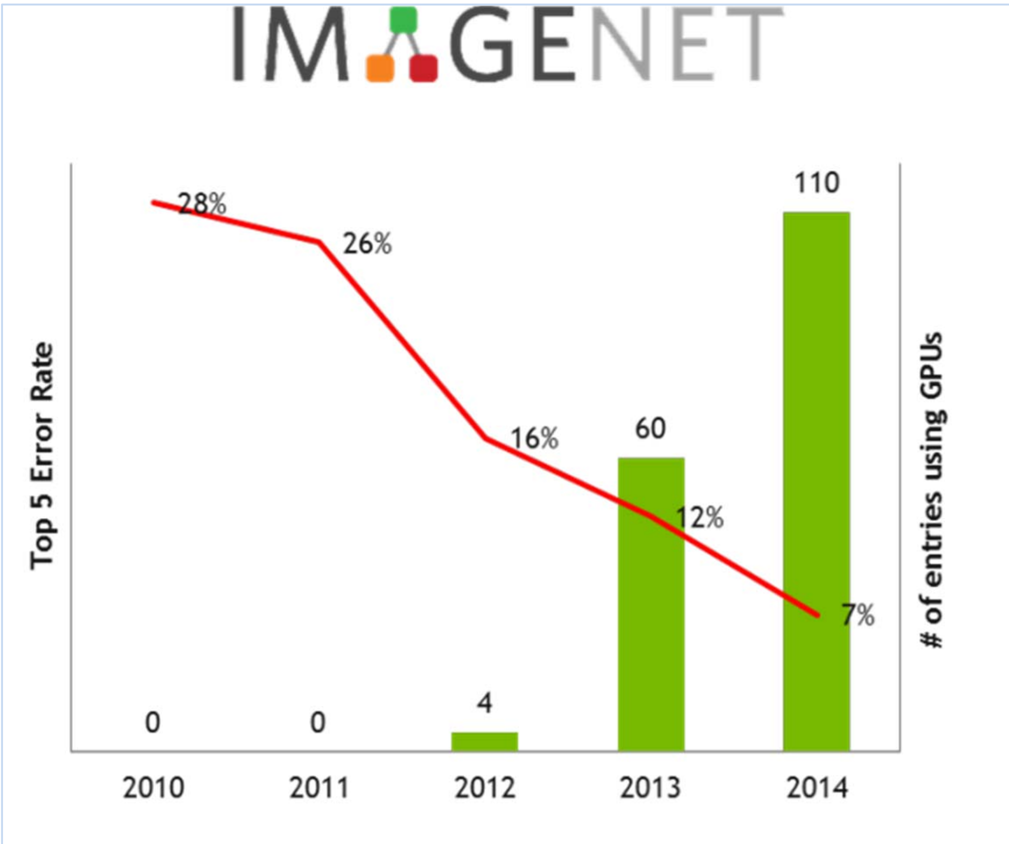
"Smart" Software Can Be Tricked into Seeing What Isn't There



Nguyen, et al, CVPR 2015

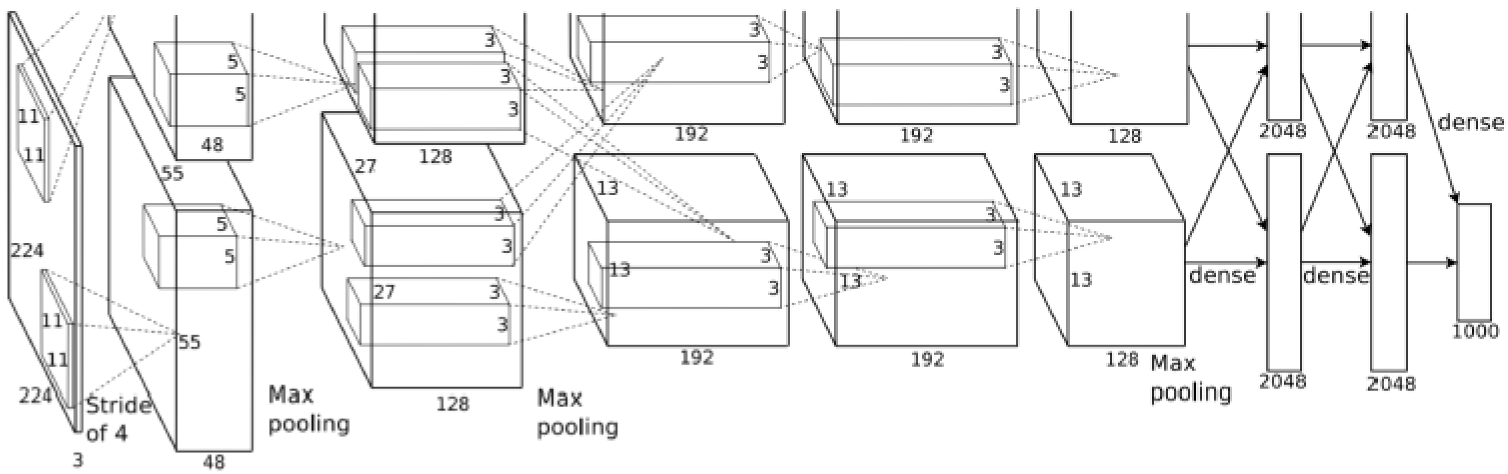
Research Issue: Compelxity

- How to handle huge data and computational cost?



Circulant Neural Networks

AlexNet (Krizhevsky, et al, 2012)



Fully connected layer in neural networks (16 millions parameters to learn !)

- Captures global information.
- 90 - 95% of the memory in AlexNet.
- 20 - 30% of the computational time.

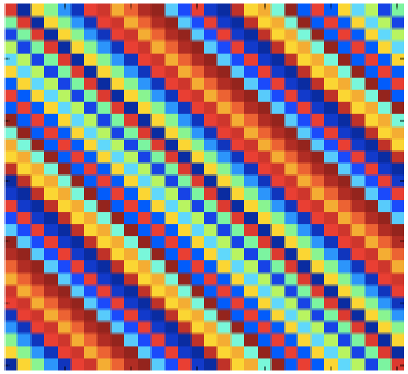
Circulant Binary Embedding (CBE)

Replace fully connected layer with structured component:

$$h(\mathbf{x}) = \phi(\mathbf{RDx}) \text{ , } \phi(\cdot)\text{: a nonlinear activation function}$$

- **R** is a **circulant matrix**, defined by $\mathbf{r} = (r_0, r_1, \cdots, r_{d-1})^T$

$$\mathbf{R} = \mathrm{circ}(\mathbf{r}) := \begin{bmatrix} r_0 & r_{d-1} & \cdots & r_2 & r_1 \\ r_1 & r_0 & r_{d-1} & & r_2 \\ \vdots & r_1 & r_0 & \ddots & \vdots \\ r_{d-2} & & \ddots & \ddots & r_{d-1} \\ r_{d-1} & r_{d-2} & \cdots & r_1 & r_0 \end{bmatrix}$$



- **D** is a diagonal matrix, each entry ± 1 with probability $1/2$ (random sign flipping, dropped to simplify notation)
- Gradient computation in back-propagation can be done using FFT.

Y. Cheng, F. Yu, R. Feris, S. Kumar, and S.-F. Chang, ICCV 2015
(and also works proposing structured efficient DL)

Circulant Neural Networks (cont'd)

Method	Time	Space	Time (Learning)
Conventional	$\mathcal{O}(d^2)$	$\mathcal{O}(d^2)$	$\mathcal{O}(ntd^2)$
Ours	$\mathcal{O}(d \log d)$	$\mathcal{O}(d)$	$\mathcal{O}(ntd \log d)$

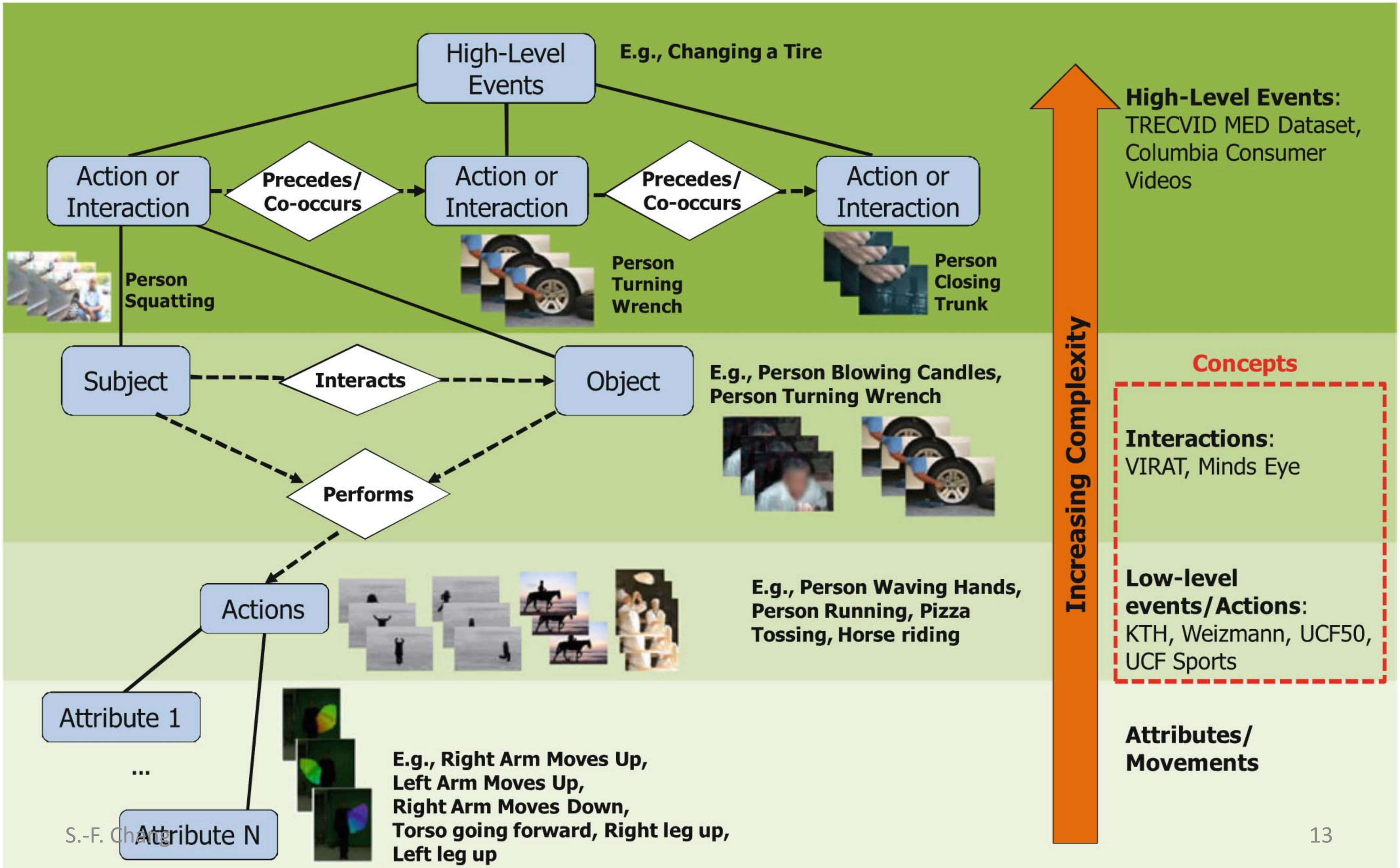
1000x less parameters in fc

- 4K $\times$ 4K layer: 64MB $\rightarrow$ 16KB

Method	Top-5 Err	Top-1 Err	Memory
AlexNet (rand)	33.5%	61.7%	233.2MB
C-AlexNet (rand)	35.2%	62.8%	20.5MB
AlexNet	17.1 %	42.8%	233.2MB
C-AlexNet	19.4 %	44.1%	20.5MB
More parameters	17.8 %	43.2%	20.7MB
Reduced AlexNet	37.2 %	65.3%	20.7MB

Open Issue: How to Describe Complex Events in Video?

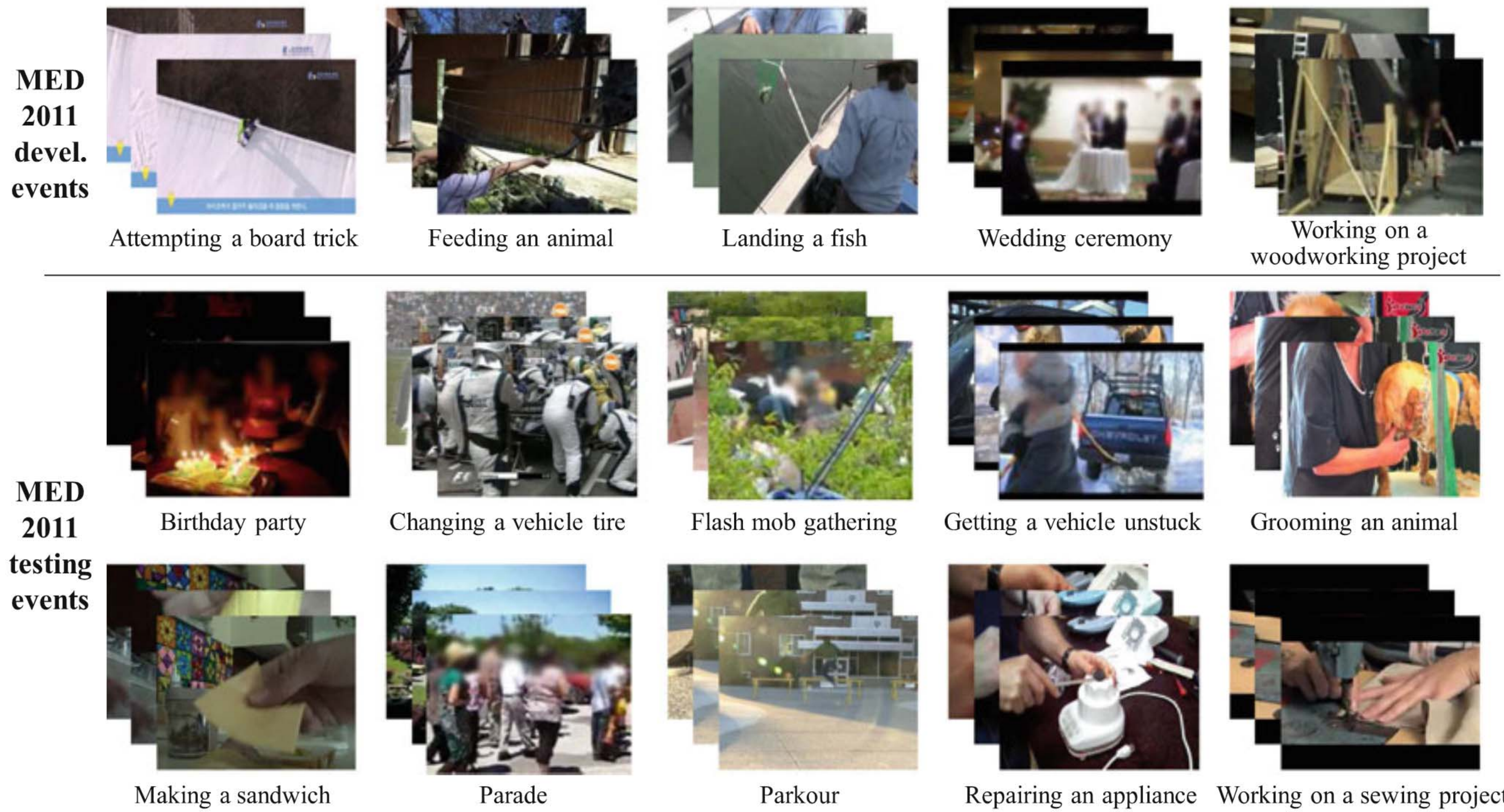
- Y. Jiang et al., high level events recognition in unconstrained videos, IJMIR, 2012 (survey)



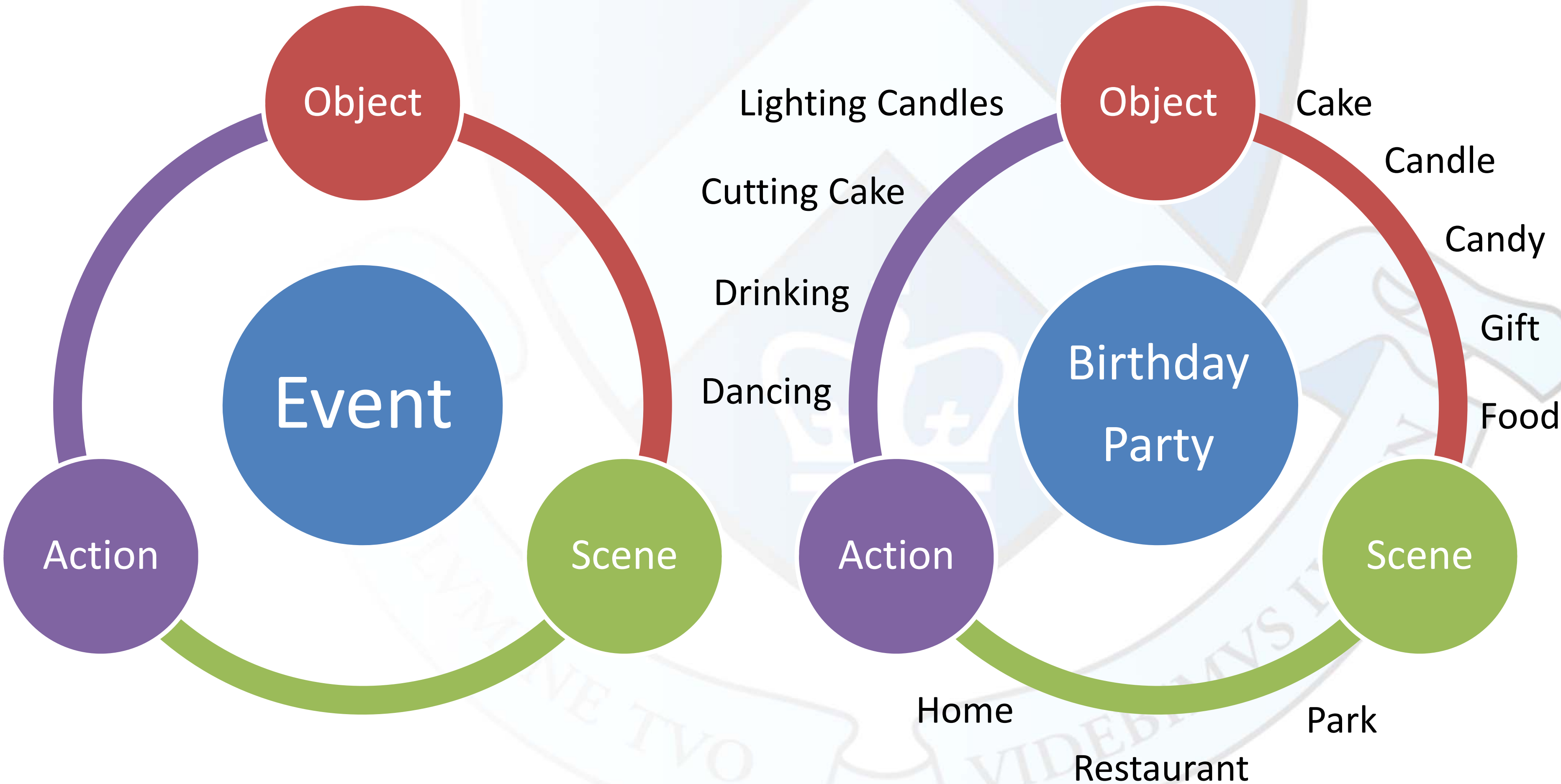
Research: Complex Video Event Recognition

K.-T. Lai; F. X. Yu; M.-S. Chen; S.-F. Chang. CVPR 2014

- Detecting complex events in ~ 100,000 videos

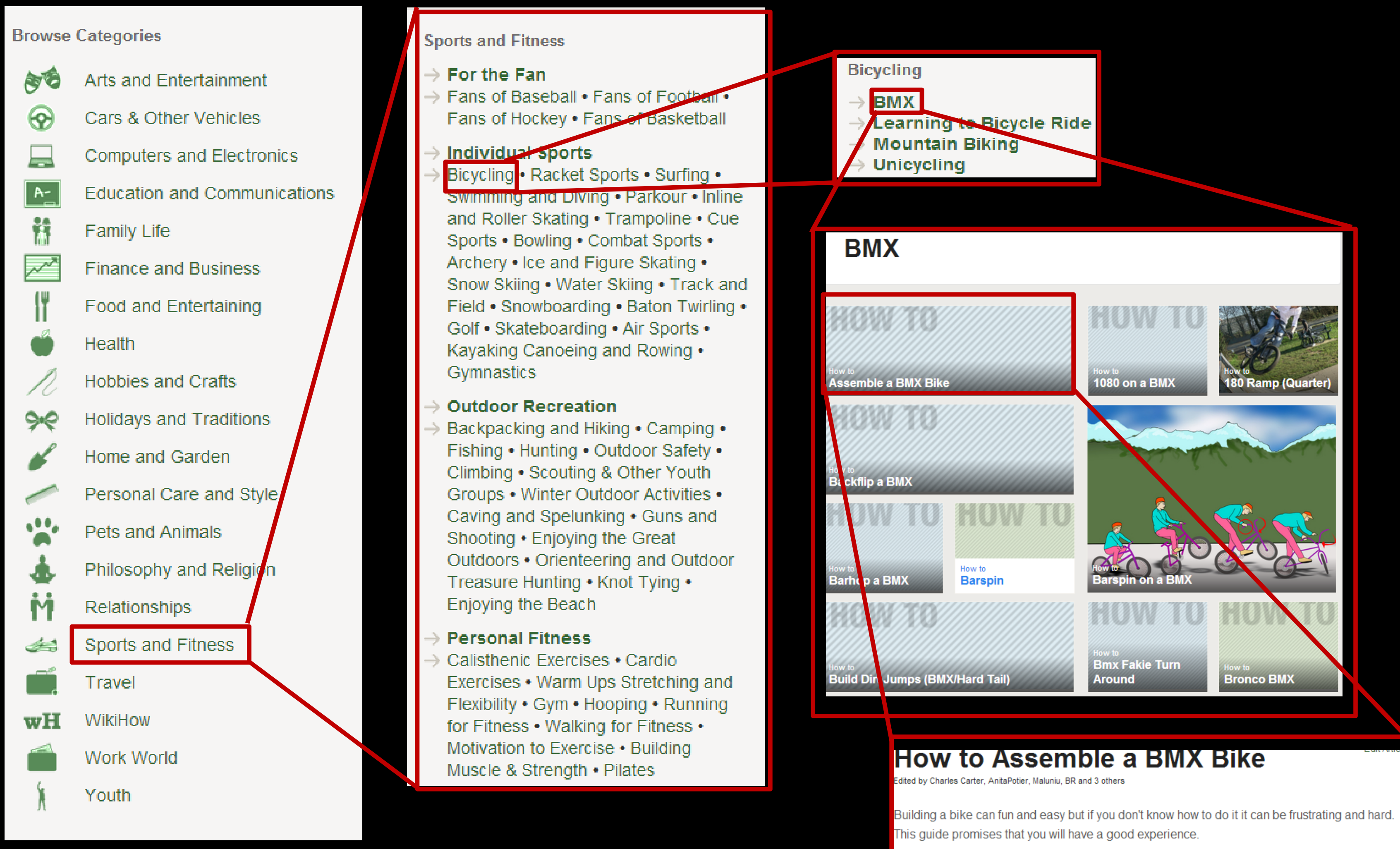


What Salient Concepts Are Associated with Each Video Event?



Inspiration from Crowd – WikiHow

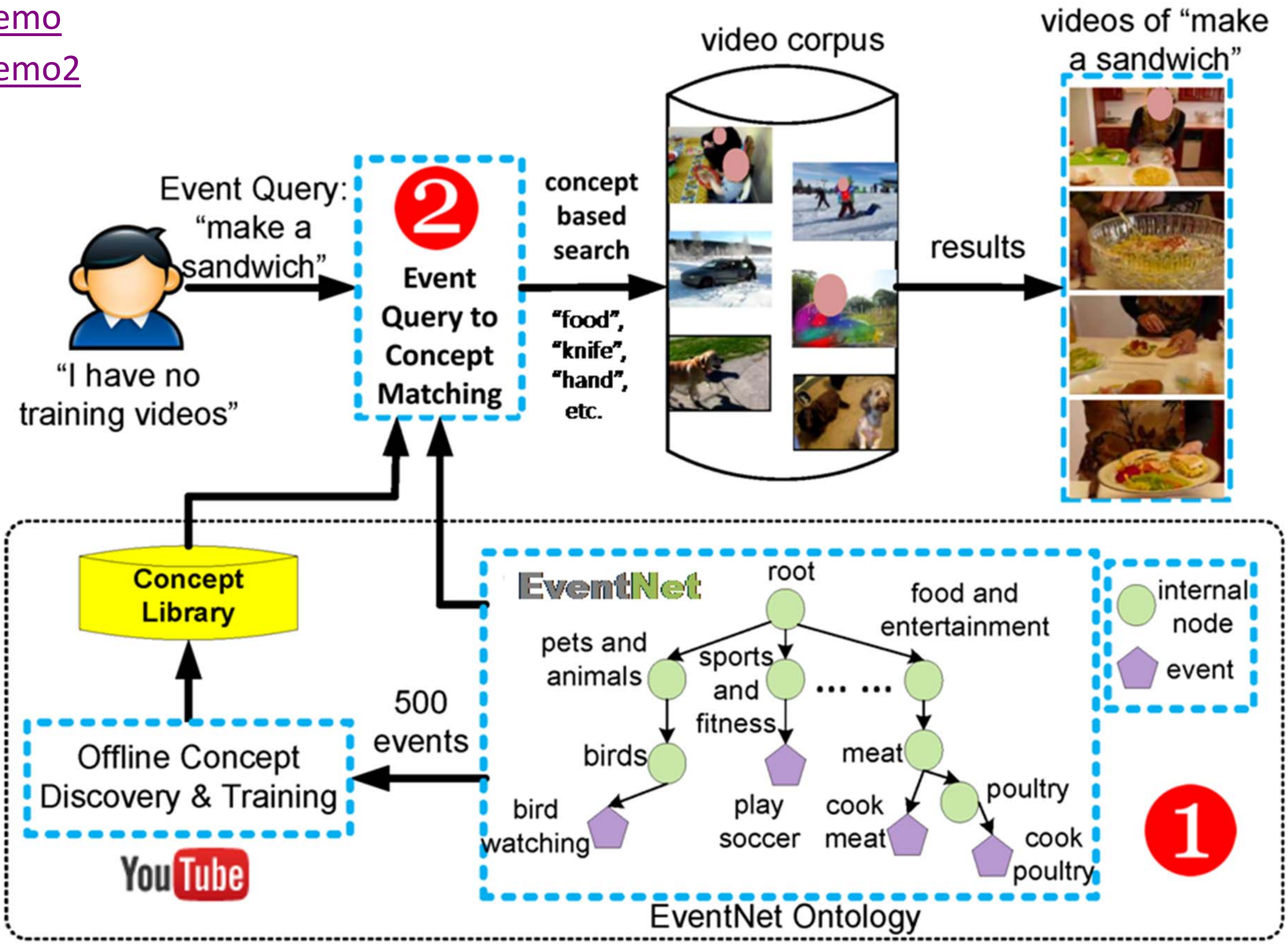
- A wiki contains ~300,000 articles on 2,803 “how to” category.
- All articles are organized into a hierarchical structure.



Columbia Video EventNet Ontology (500 events, 4,800 concepts)

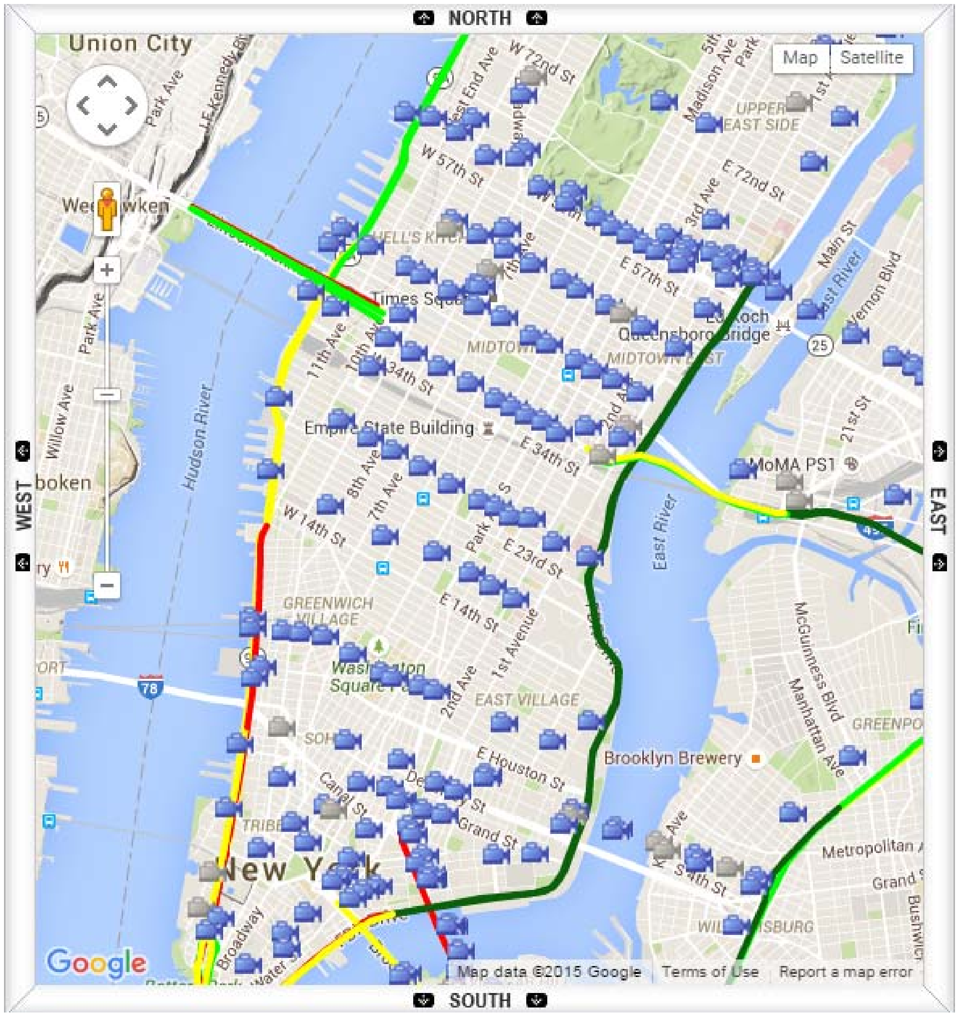
Demo

Demo2

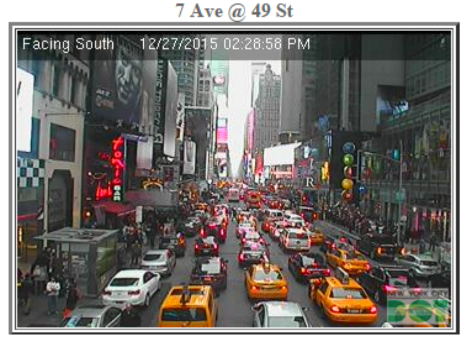


Ye, et al, EventNet, ACMMM '15, eventnet.ee.columbia.edu

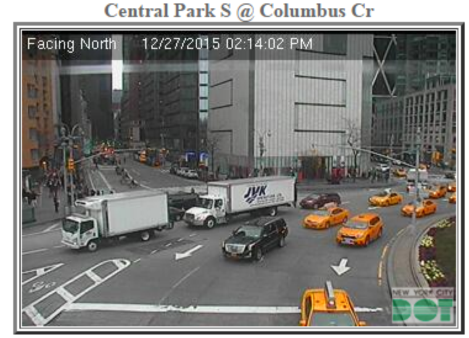
Video in Large Data System: Traffic Cams



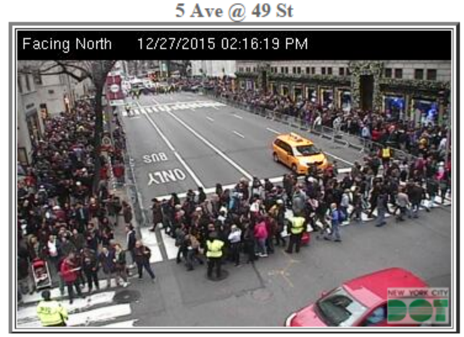
7 Ave @ 49 St



Central Park S @ Columbus Cr

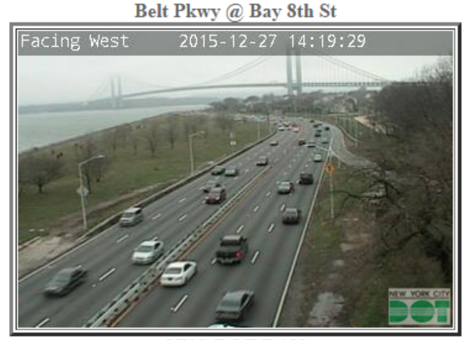


5 Ave @ 49 St



Broadway @ 43 St

Belt Pkwy @ Bay 8th St



NYC DOT RAN

MHB-31 Bklyn N URW @ Anch

NYC DOT SEM

Urban Sensing: City webcam and social media

- 400+ traffic webcams in NYC: http://nyctmc.org/
- Resolution 352x240. Framerate between 1fps and 1 frame every 3s.
- Data-agreement: http://www.nyc.gov/html/dot/downloads/pdf/video-partnership-agreement.pdf
- No ground truth. Event calendars from OpenData and DOT API
    - https://developer.cityofnewyork.us/api/dot-data-feeds-beta
    - https://developer.cityofnewyork.us/api/events-calendar
    - http://www.nyc.gov/html/dot/html/motorist/wkndtraf.shtml

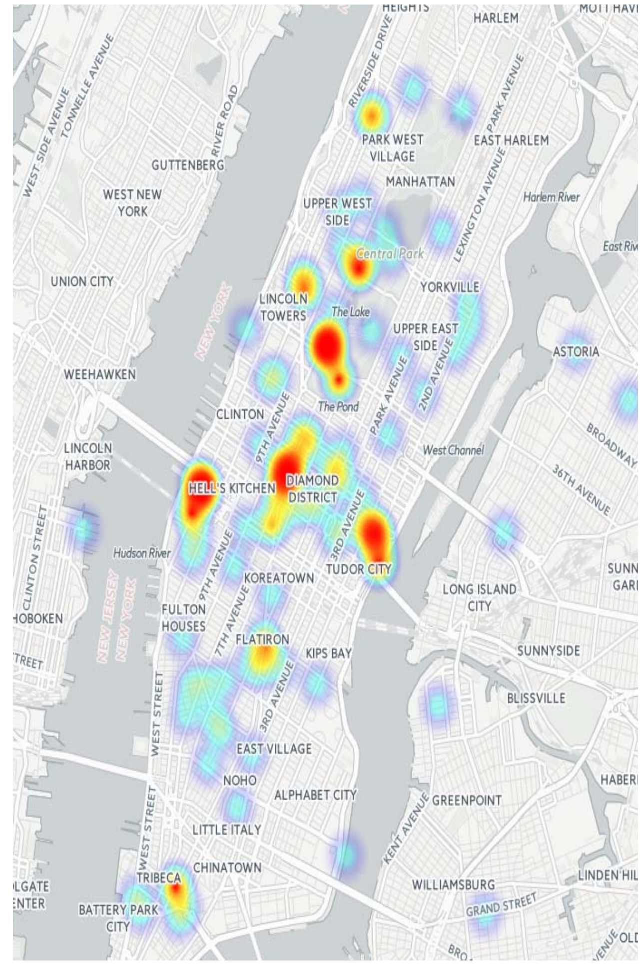
event	*date*	*time*	*location*
CBGB Music Festival	12 Oct	10am-7pm	Broadway 51 Street
Hispanic Parade	12 Oct	12pm-5pm	5th Avenue
Columbus Day Parade	13 Oct	11am-5pm	5th Avenue
Saint Patrick's Day Parade	17 Mar	12pm-5pm	5th Avenue
Million March NYC Protest	13 Dec	2pm-5pm	Washington Square Park, 5th Avenue, Foley Square

Example Tweets (Wu & Kankanhalli, WWW’15)

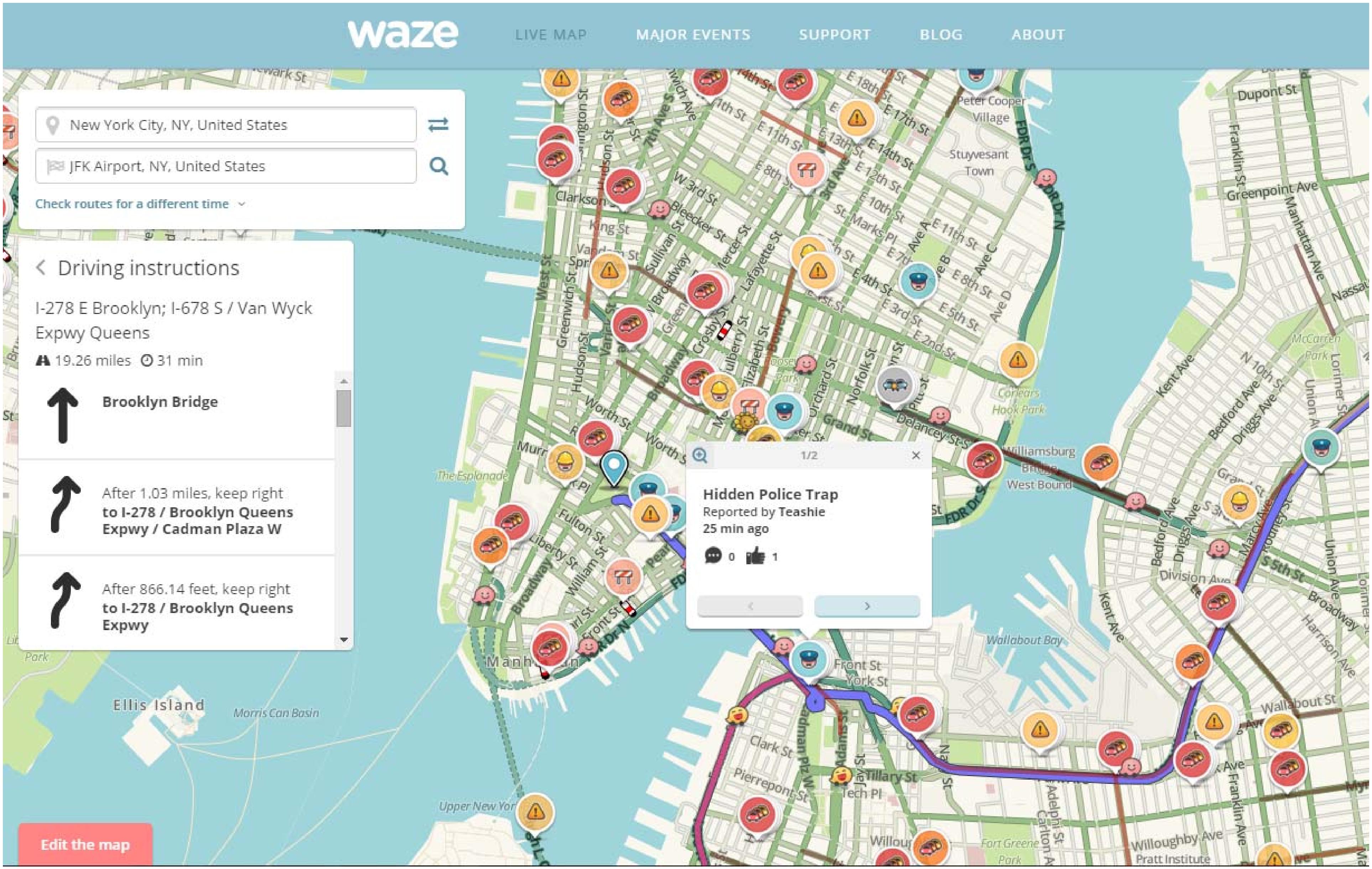
2011 2015-10-3105:45:21 Binnen... Incl honered guest bandje ? 1 day 2 #tcsnycmarathon @ Times Square, New York City https://t.co/ew4buSUhDR 40.758895 -73.985131

2074 2015-10-3106:07:34 The honey and I got new running shoes at the NYC marathon expo!! I have never been in a more¡ https://t.co/UIeMSZhZdy 40.75759585 -74.00118993

2226 2015-10-3106:55:00 Not a bad view from my hotel room watching the sun come up!! Off to the marathon expo to collect my¡ https://t.co/HWr9OH88Ht 40.75807569 -73.9753008

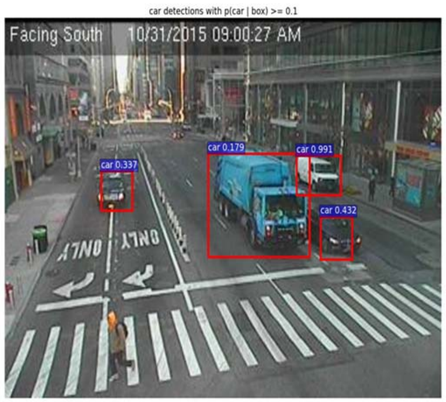


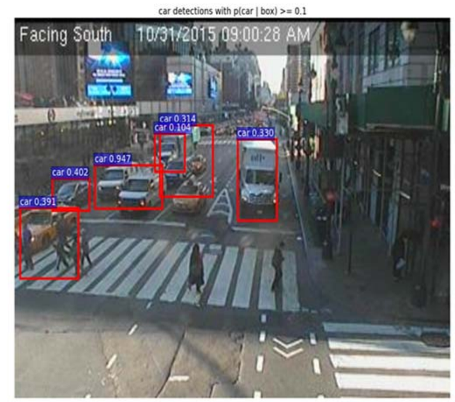
Social Crowd Reporting (e.g., Waze)



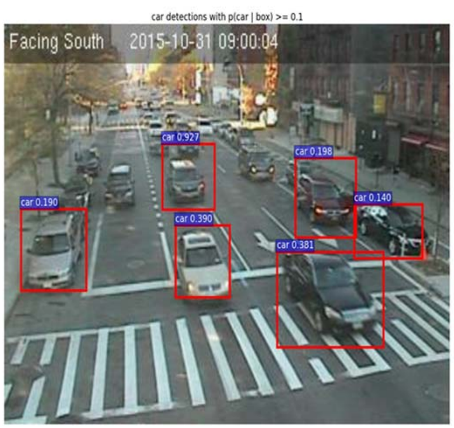
Computer Vision Tools

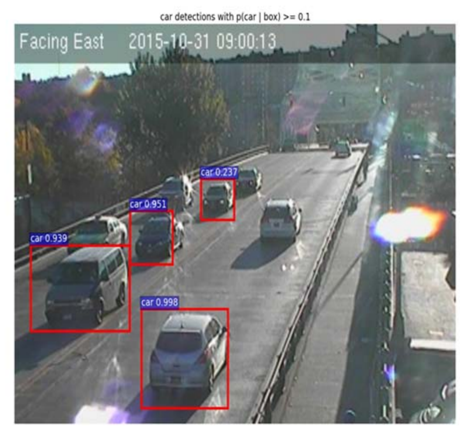
car detection - success in most cases:





car detection - some missing detections:





Potential Applications and on-going research:

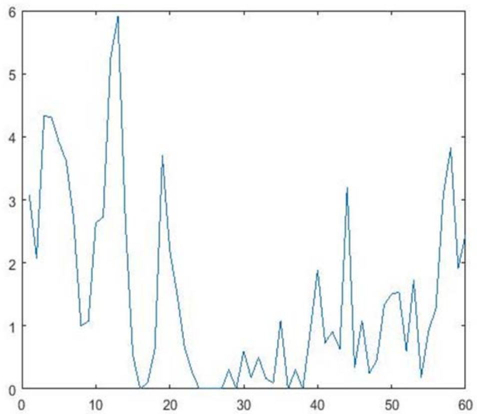
- Traffic patterns
- Accidents
- Individual vehicle condition:
  aggressive driving, abnormal driving, car breakdown
- Vehicle tracking across cameras

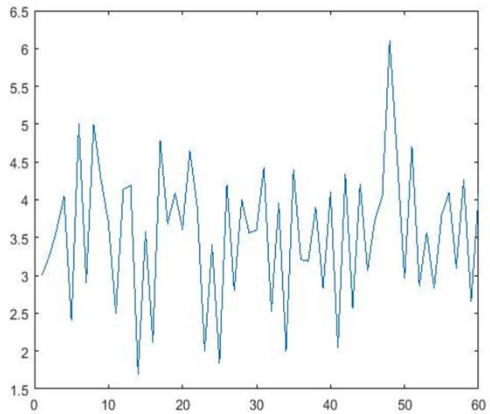
observations

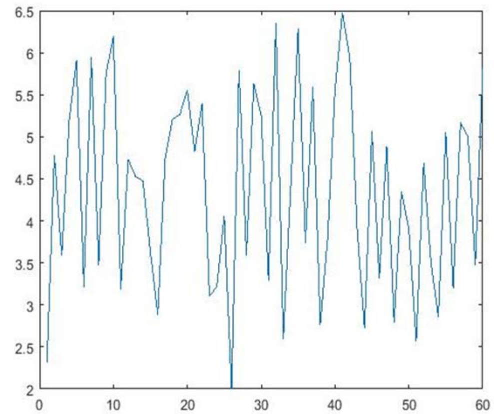
Three nearby cameras:

157-92_Holland_Tunnel
430-320_RT._9A_@_N._Moore_St
715-664_West_St_@_West_Houston_St

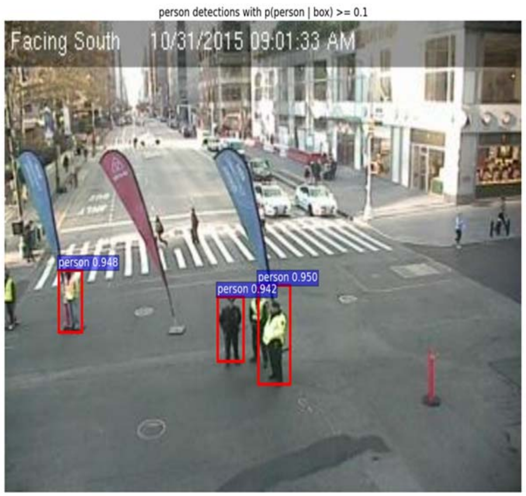
Number of cars detected on each image. 8am~9am. Dec 3rd. y: #car, x: minute.

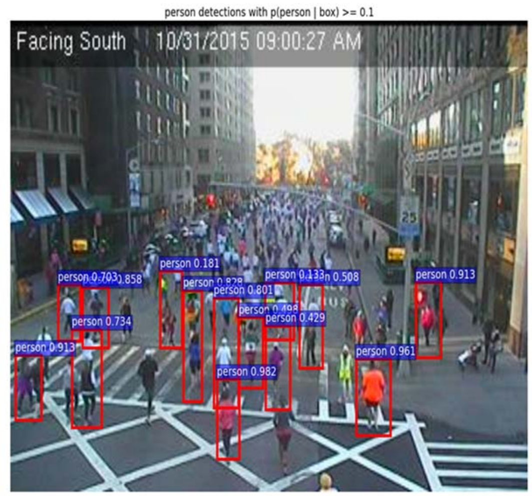






Person Detection



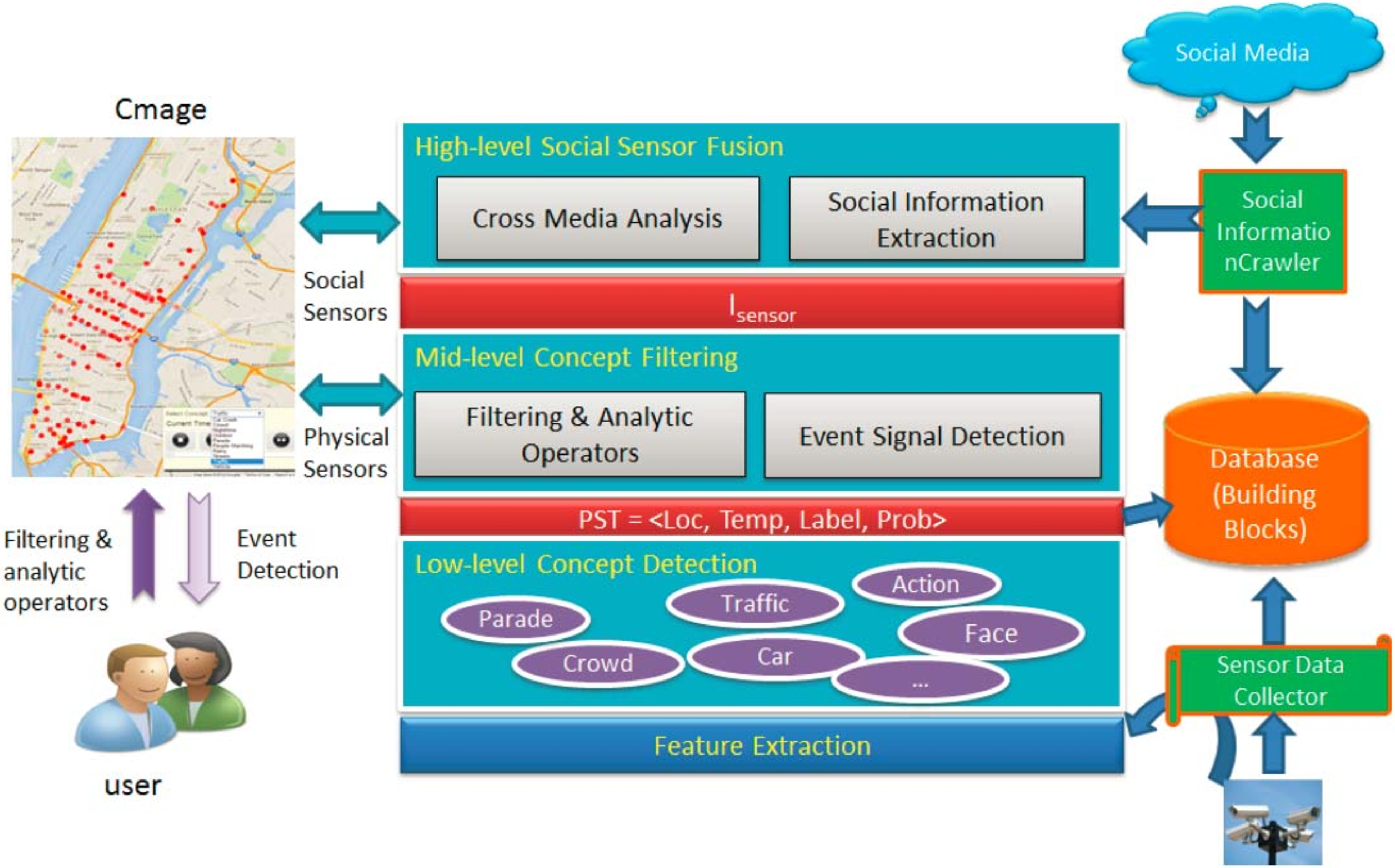


Possible Applications:

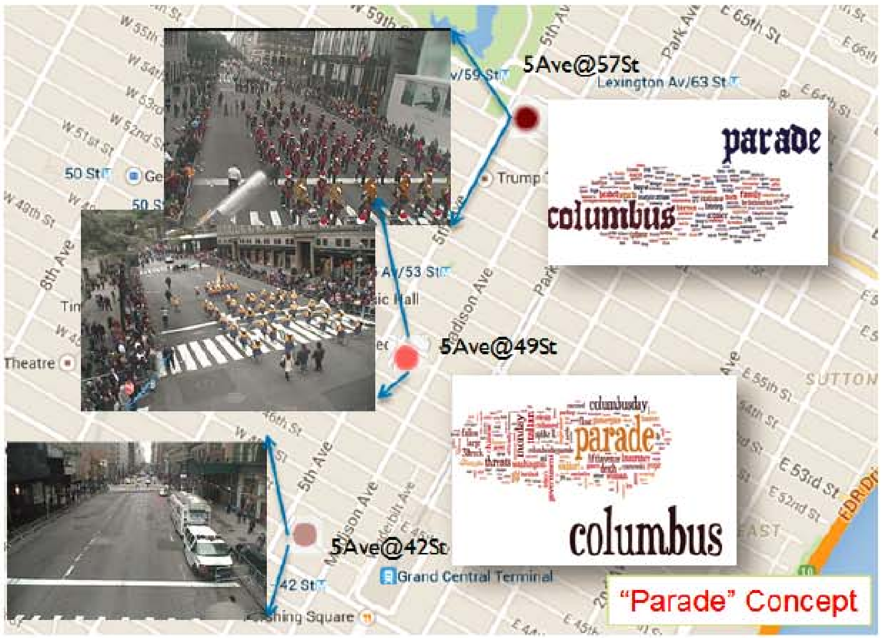
- People density: few people, crowd, crowd motion
- Categorize events: parade, marathon
- Social Behavior: random crowd vs. affiliated groups

DEMO

Social Tweeting Camera, (Wang & Kankanhalli, WWW'15)



Combining Camera Sensors and Social Signals



(a) Social Sensor Fusion for "Columbus Day Parade"

(b) Social Sensor Fusion for "Hispanic Parade" Event

Wu & Kankanhalli, WWW'15

Ongoing research Issues

Deeper Analysis

- Vision: people, object, setting, events
- NLP: entity, topic, events, sentiments

Joint Analysis

- Naming events seen in video
- Illustrating entities mentioned in text
- Disambiguate concepts (e.g., “truck”, “sign”)

Social Sharing -> Opinion Expression

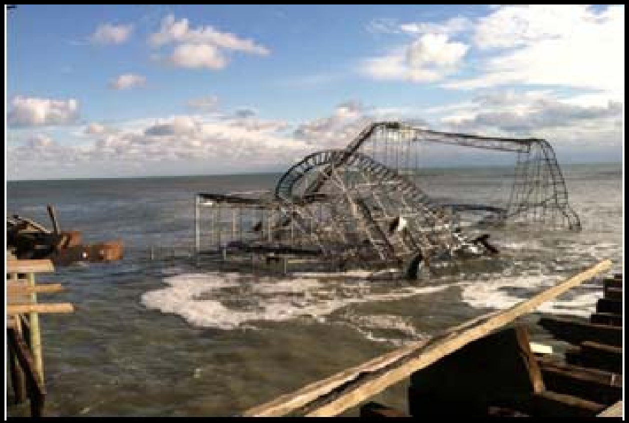
(socialdefender.com)

The Power of Social (Visual) Multimedia

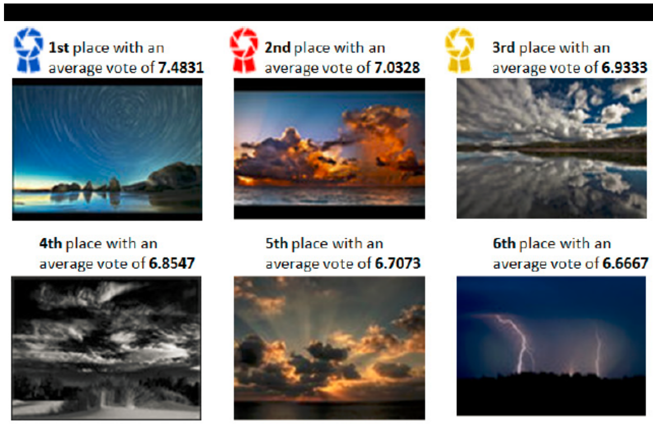
2012 Tweets of the Year

@BarackObama: Four more years.

@Brynn4NY: Rollercoaster at sea.



@Fang-Ru: Queen of the far far away land.



(CVPR 2014 Tutorial)

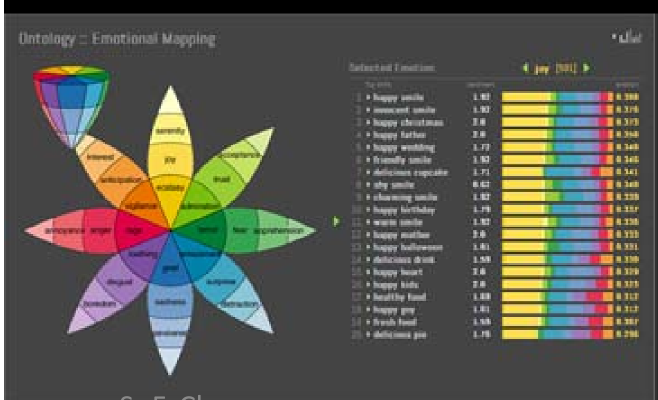
Aesthetics

Interestingness

Beyond Semantics

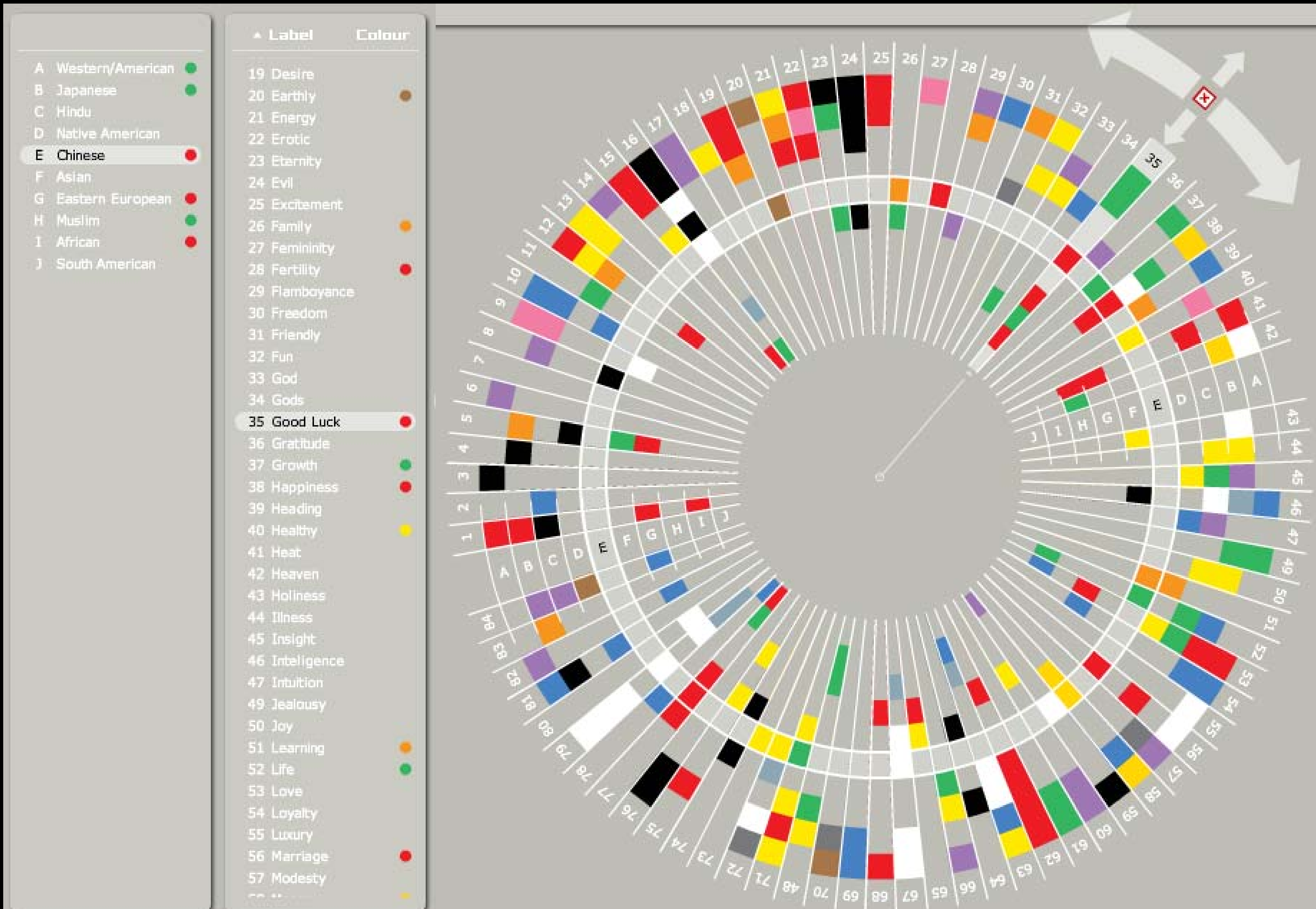
Emotion

Style



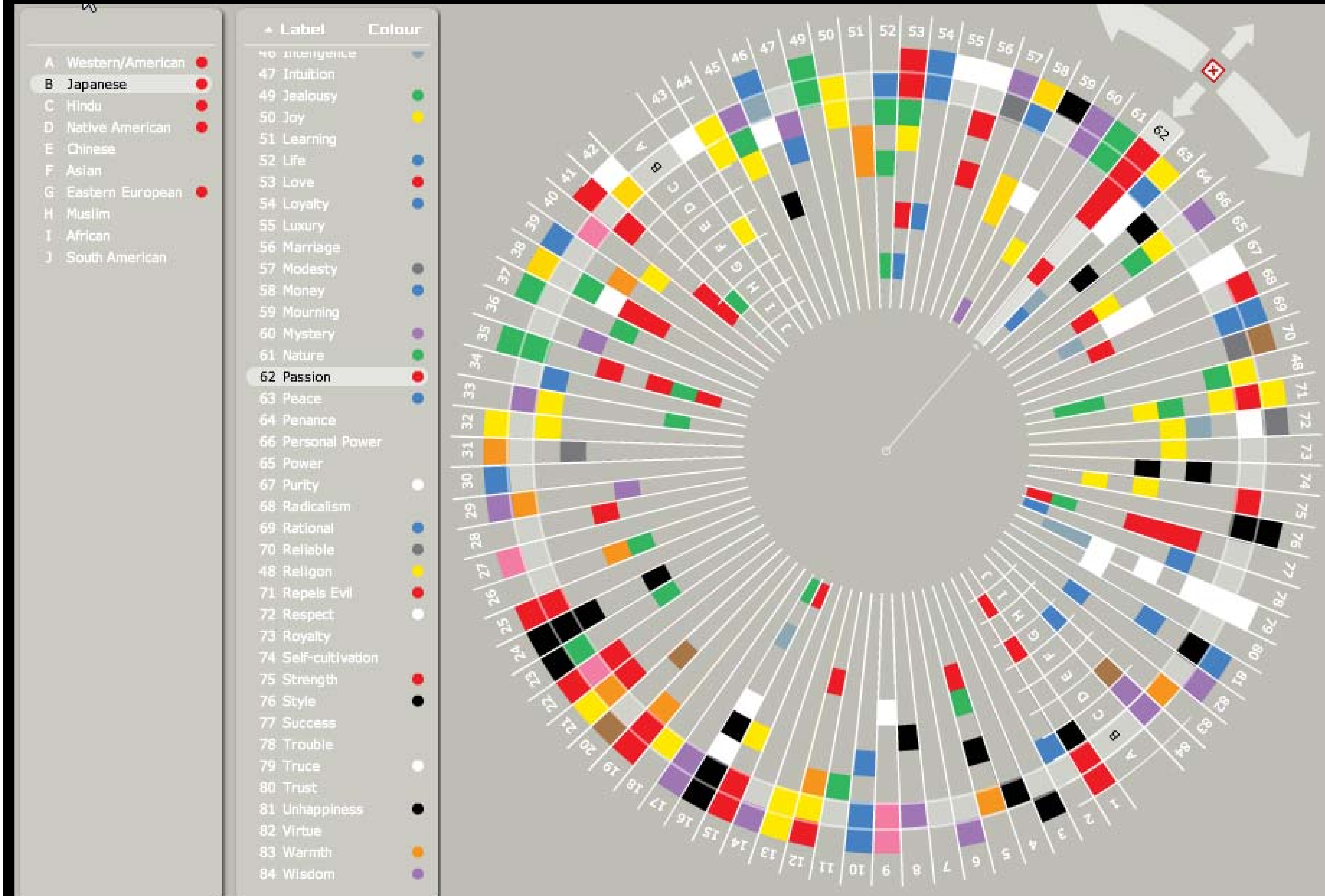
Others:, Creativity, Intent, Memorable …

Culture plays an important role in visual presentation



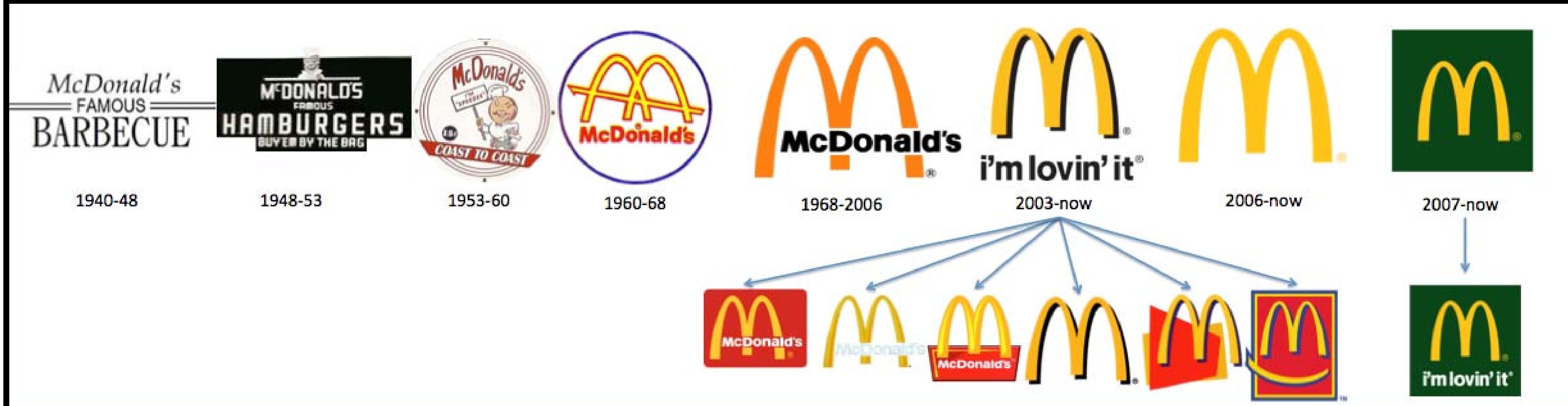
David McCandless , Information is Beautiful, Interactive tools by Lab Zoho

Some colors are pan-cultural



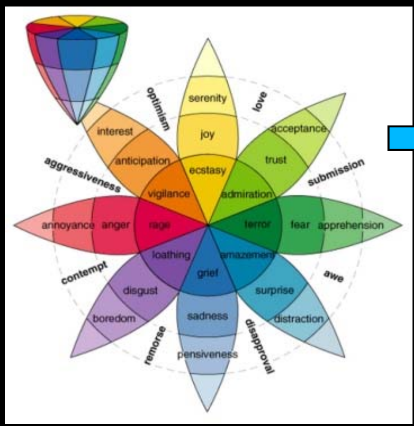
David McCandless , Information is Beautiful, Interactive tools by Lab Zoho

Color Logos in the World

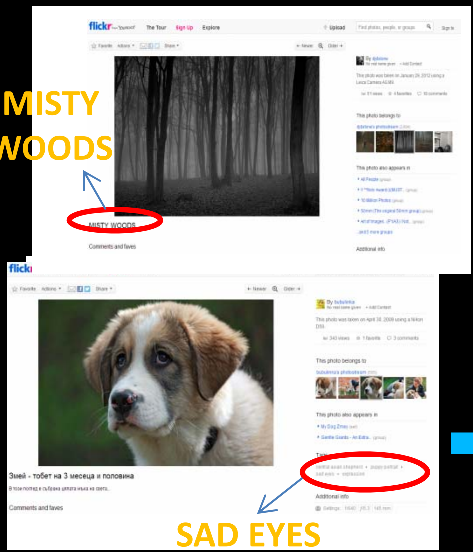


Columbia Visual Sentiment Ontology:

- Discover popular visual concepts used to express emotions



Psychology emotion wheel (8 emotions) Robert Plutchik, '91



Analyze tags with strong sentiments

Select Concepts

Build Sentiment Ontology

Borth, Ji, Chen, Breuel, Chang, *Large-Scale Visual Sentiment Ontology*, ACM Multimedia 2013

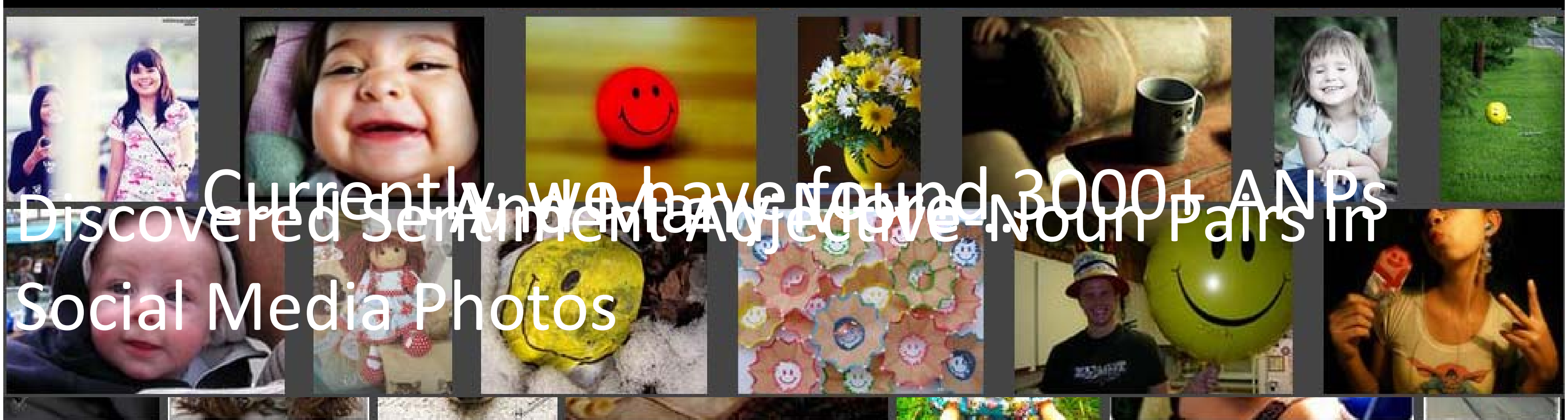
Frequent Photo Tags Related to Emotions

S.-F. Chang

Discovered Sentiment Adjective-Noun Pairs in Social Media Photos

Currently, we have found 3000+ ANPs

And Many More…

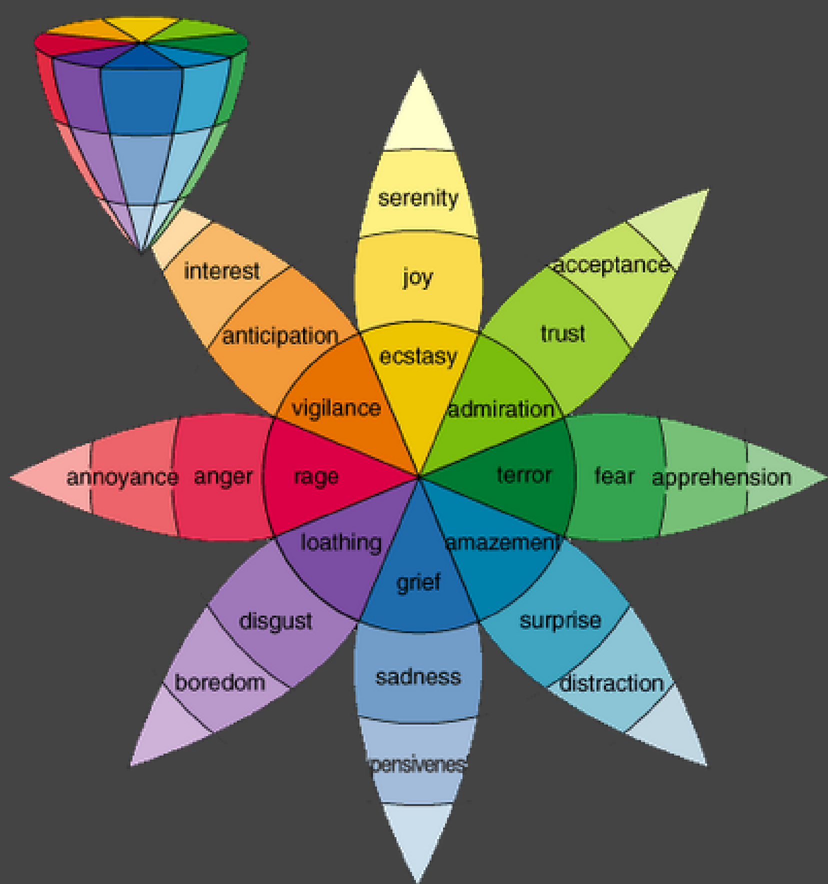


Visual Sentiment Ontology (Browser)

Home :: Ontology :: Adjective Noun Pairs :: Downloads :: About

Visual Sentiment Ontology

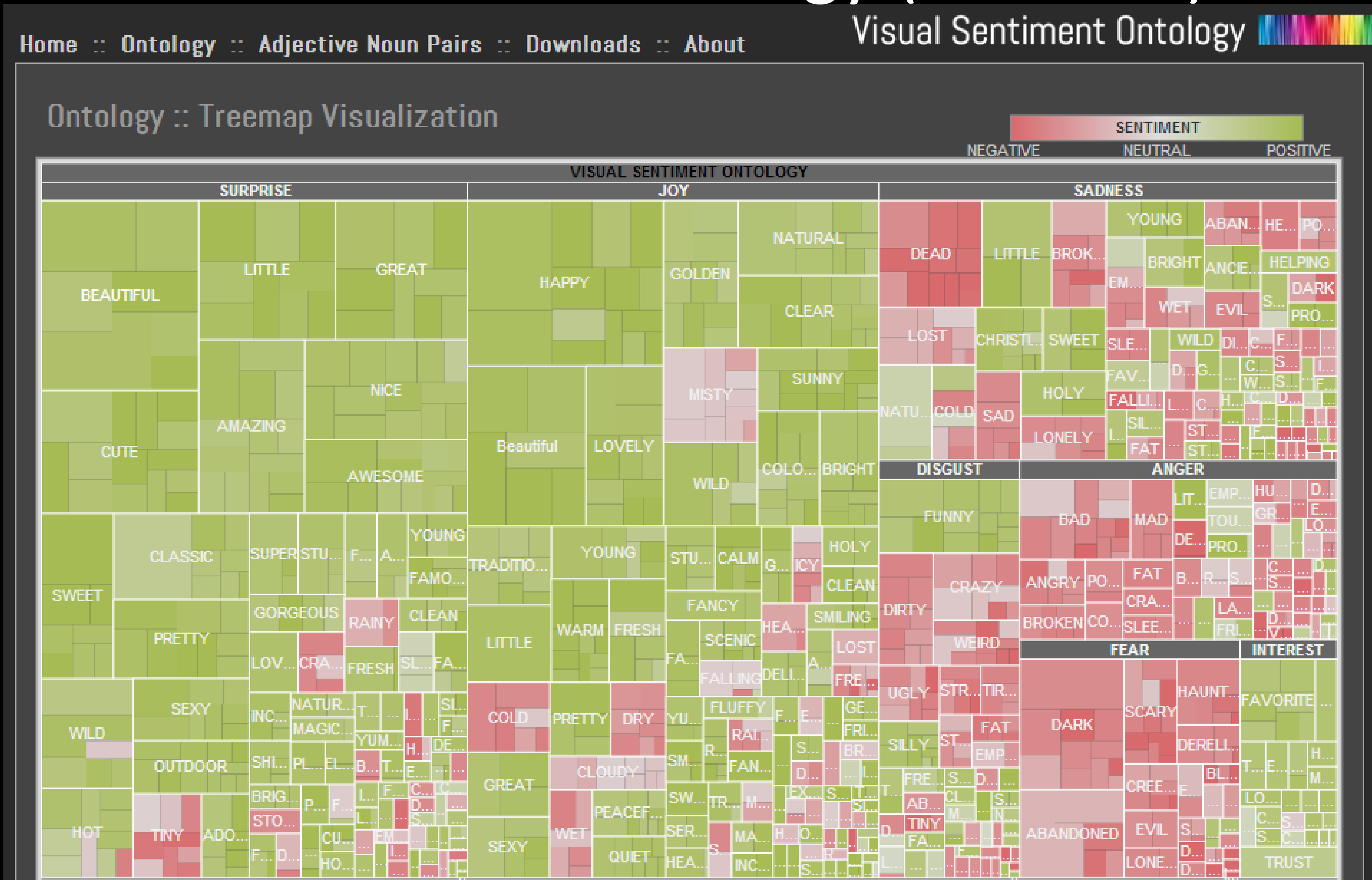
Ontology :: Emotional Mapping



Selected Emotion: joy [591]

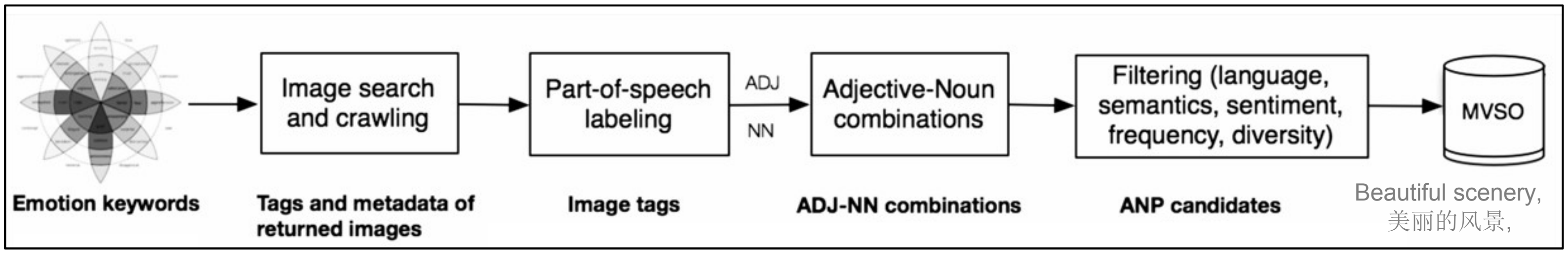
	Top ANPs	sentiment	emotion
1.	happy smile	1.92	0.388
2.	innocent smile	1.92	0.376
3.	happy christmas	2.0	0.373
4.	happy father	2.0	0.358
5.	happy wedding	1.72	0.348
6.	friendly smile	1.92	0.346
7.	delicious cupcake	1.71	0.341
8.	shy smile	0.62	0.340
9.	charming smile	1.92	0.339
10.	happy birthday	1.79	0.337
11.	warm smile	1.92	0.336
12.	happy mother	2.0	0.333
13.	happy halloween	1.81	0.331
14.	delicious drink	1.59	0.330
15.	happy heart	2.0	0.329
16.	happy kids	2.0	0.323
17.	healthy food	1.69	0.312
18.	happy guy	1.61	0.312
19.	fresh food	1.59	0.307
20.	delicious pie	1.76	0.296

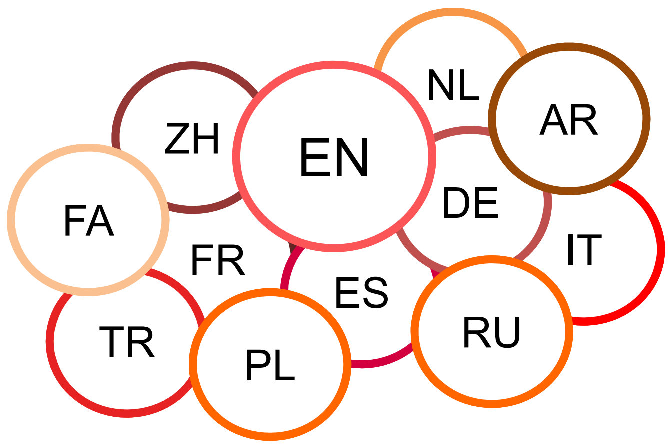
Visual Sentiment Ontology (Browser)



THE MULTILINGUAL VISUAL SENTIMENT ONTOLOGY

MVSO Pipeline





16,000 visual concepts, 7 million images, 12 languages

Brendan Jou, Tao Chen, Nikolaos Pappas, Miriam Redi, Mercan Topkara, Shih-Fu Chang.
"Visual Affect Around the World: A Large-scale Multilingual Visual Sentiment Ontology," ACM Multimedia 2015.

Images + Emotions in Different Languages

(i) **english** (top: interest, joy, trust)

(ii) **spanish** (top: alegría, dolor, sorpresa)

(iii) **french** (top: surprise, joie, intérêt)

(iv) **chinese** (top: 恐怖, 歡樂, 无聊)

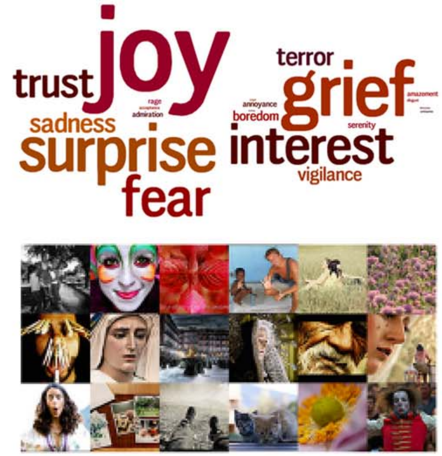
(v) **german** (top: Angst, Freude, Interesse)


(vi) **arabic** (top: الحزن, حزن, الفرح)

Emotion keyword clouds (in English)

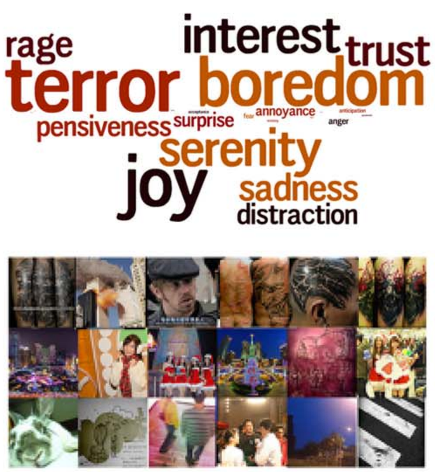
- emotion queries performed in Flickr after translating 24 emotions words from English to 15 other languages
- kept languages which have at least 100k results (search on full metadata: title, description and tags)
- word size is proportional to the number of flickr results

(i) **english** (top: interest, joy, trust)

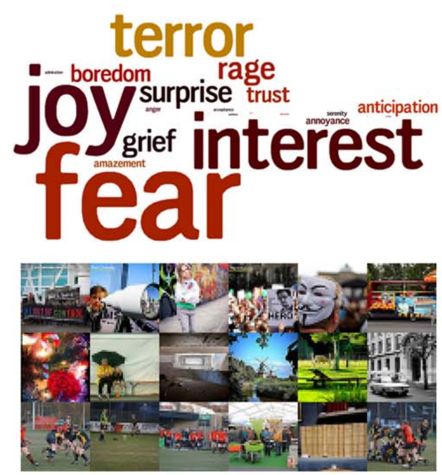


(ii) **spanish** (top: joy, grief, surprise)

(iii) **french** (top: surprise, joy, interest)



(iv) **chinese** (top: terror, joy, boredom)



(v) **german** (top: fear, joy, interest)

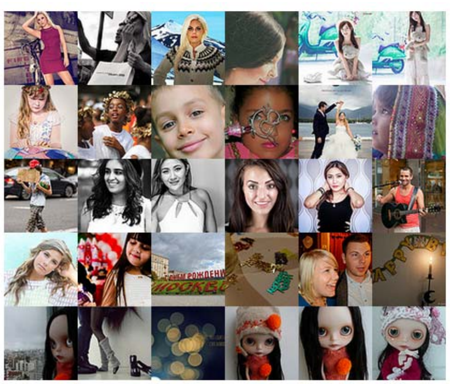
(vi) **arabic** (top: sadness, grief, joy)

Multilingual Emotion-Related Concepts for ‘joy’

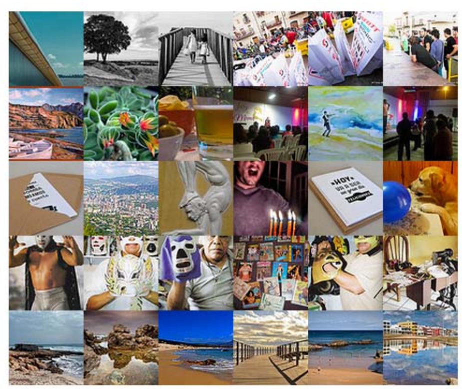
- **<u>filtering rules</u>: non-neutral sentiment, 100 exact matches on Flickr, sorted by co-occurrence with ‘joy’**

	English (joy)	**Spanish (alegría)**	**French (joie)**
# 1	beautiful girl	aire libre (open air)	bonne humeur (good mood)
# 2	happy child	feliz día (happy day)	plein air (open air)
#3	beautiful smile	feliz cumpleaños (happy birthday)	belle plage (beautiful beach)
#4	happy birthday	*libre lucha* (free fight/wrestling)	bonne année (happy new year)
#5	cute love	buen tiempo (good time)	belle femme (beautiful woman)
#6	beautiful game	natural belleza (natural beauty)	belle fille (beautiful girl)
#7	happy smile	grande agua (great water)	*parcours ludique* (fun trail)
#8	happy living	simple vida (simple life)	quotidienne vie (everyday life)
#9	beautiful portrait	libre mundo (free world)	grand soleil (great sun)
#10	beautiful man	mejores amigas (best friends)	belle fleur (beautiful flower)

#1
#2
#3
#4
#5



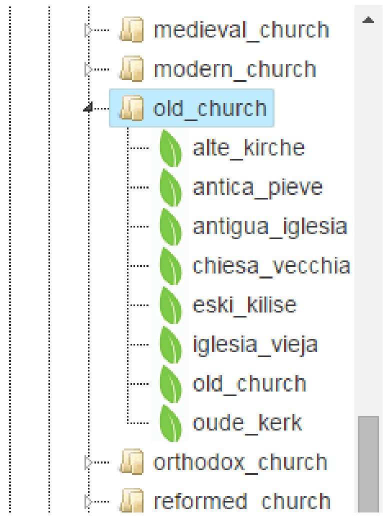
#1
#2
#3
#4
#5

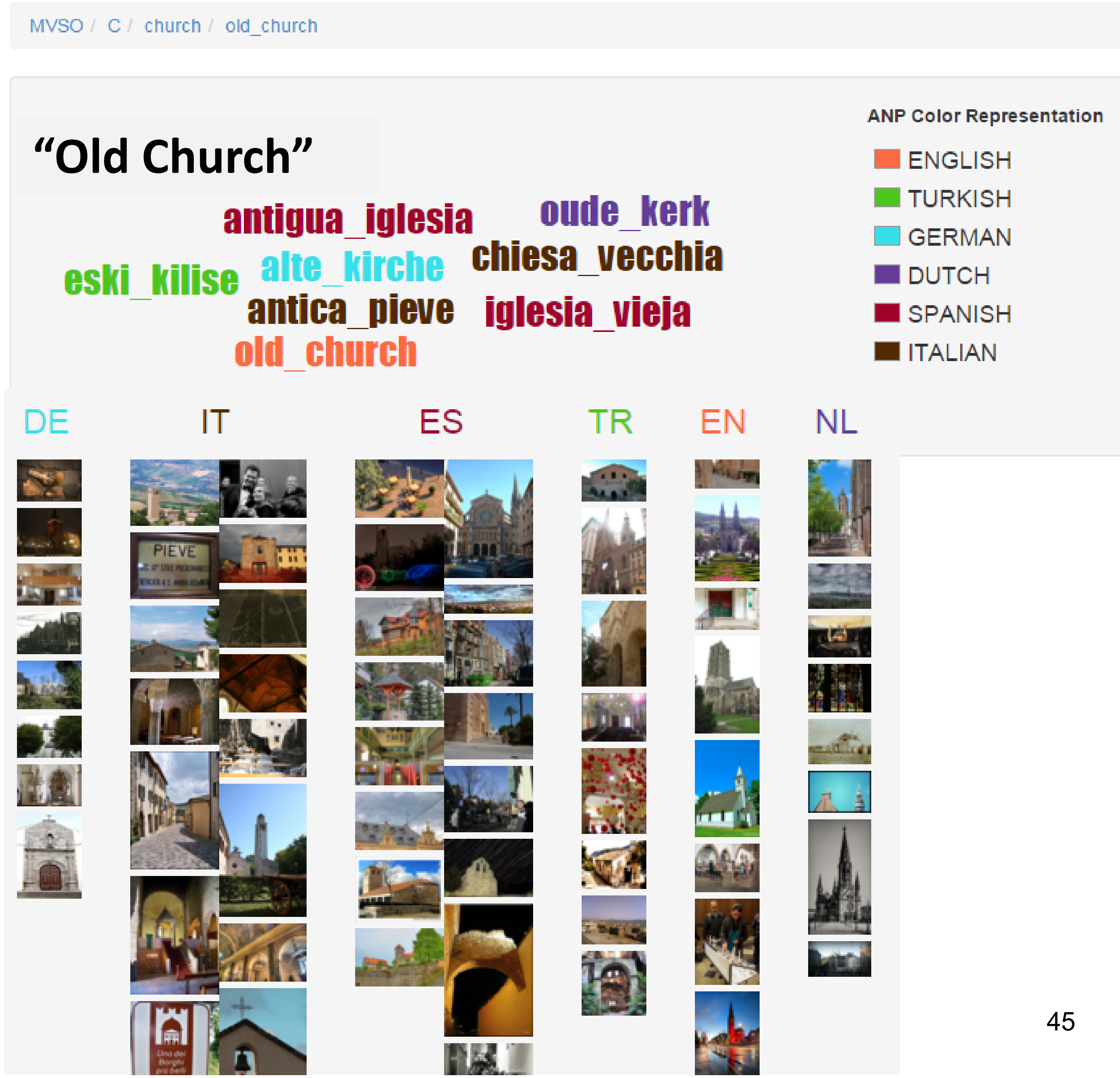


#1
#2
#3
#4
#5

How do images of the same concept vary in different languages?

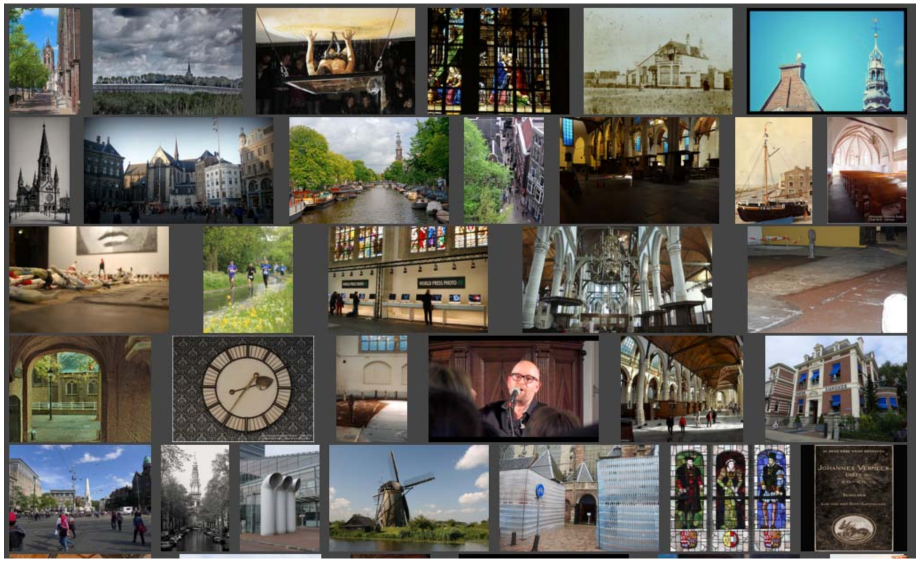
Ontology Structure



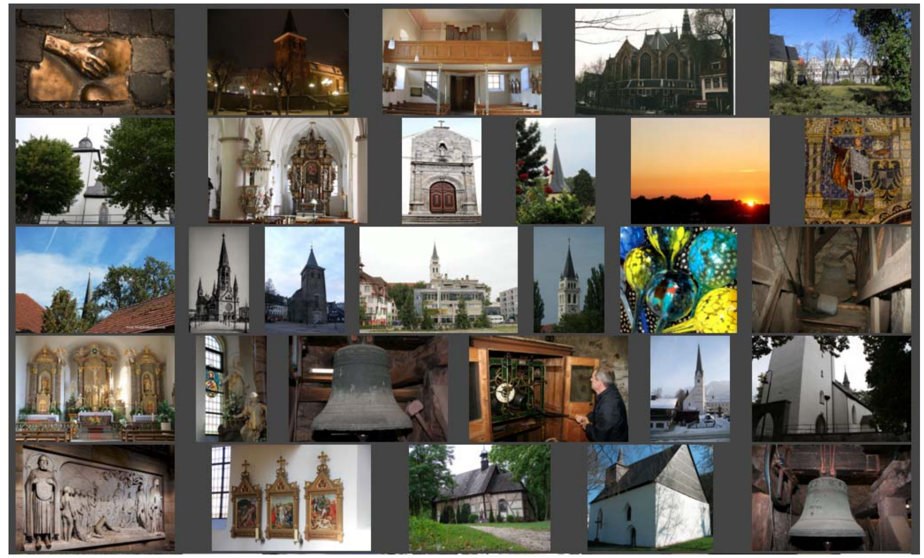


“Old Church” Images: Cultural Influences?

Dutch

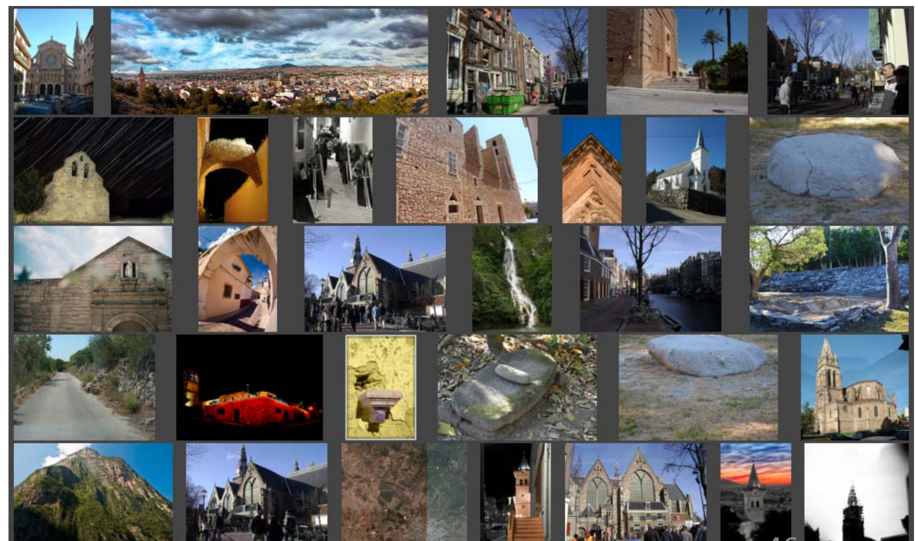


German

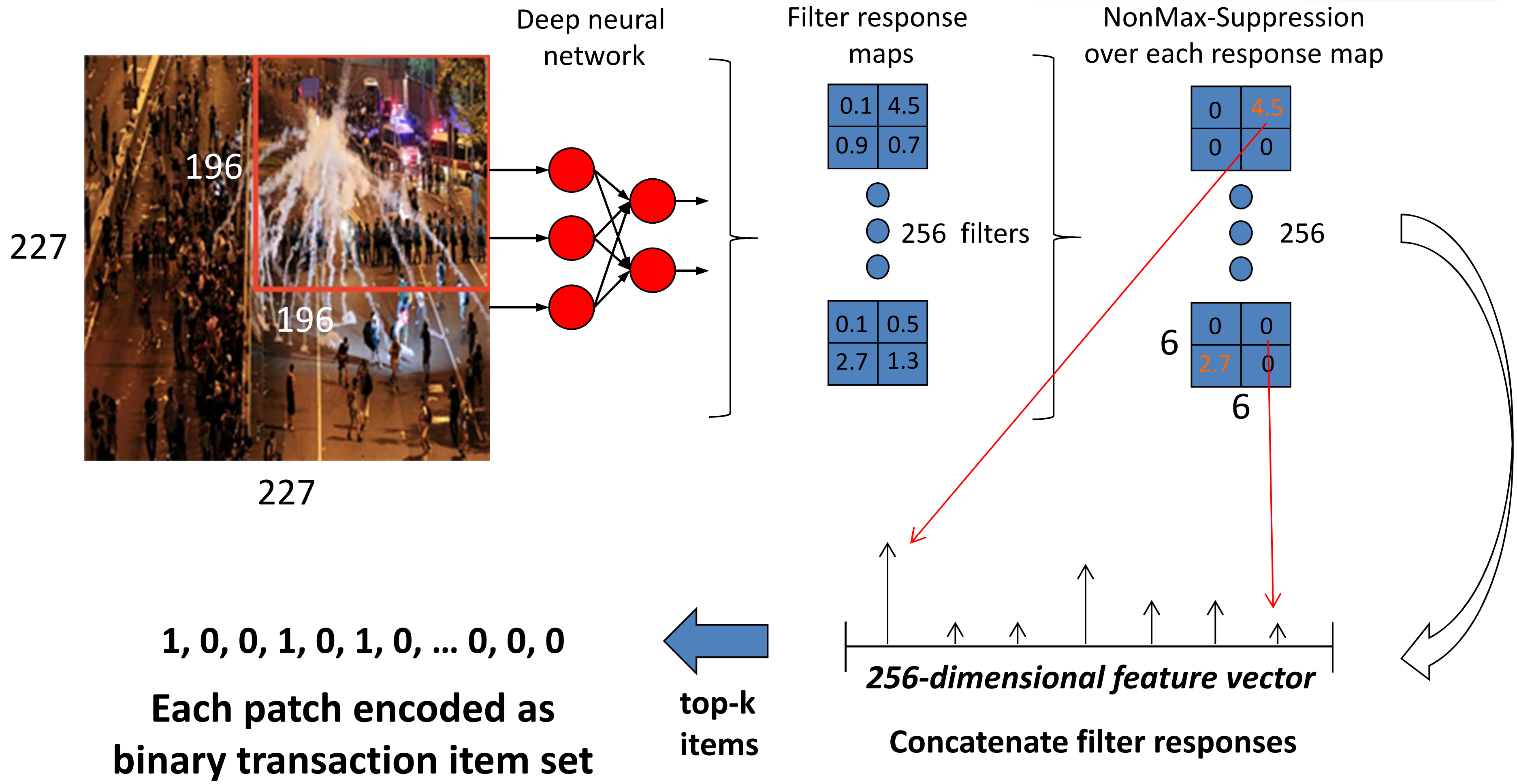


Italian

Spanish

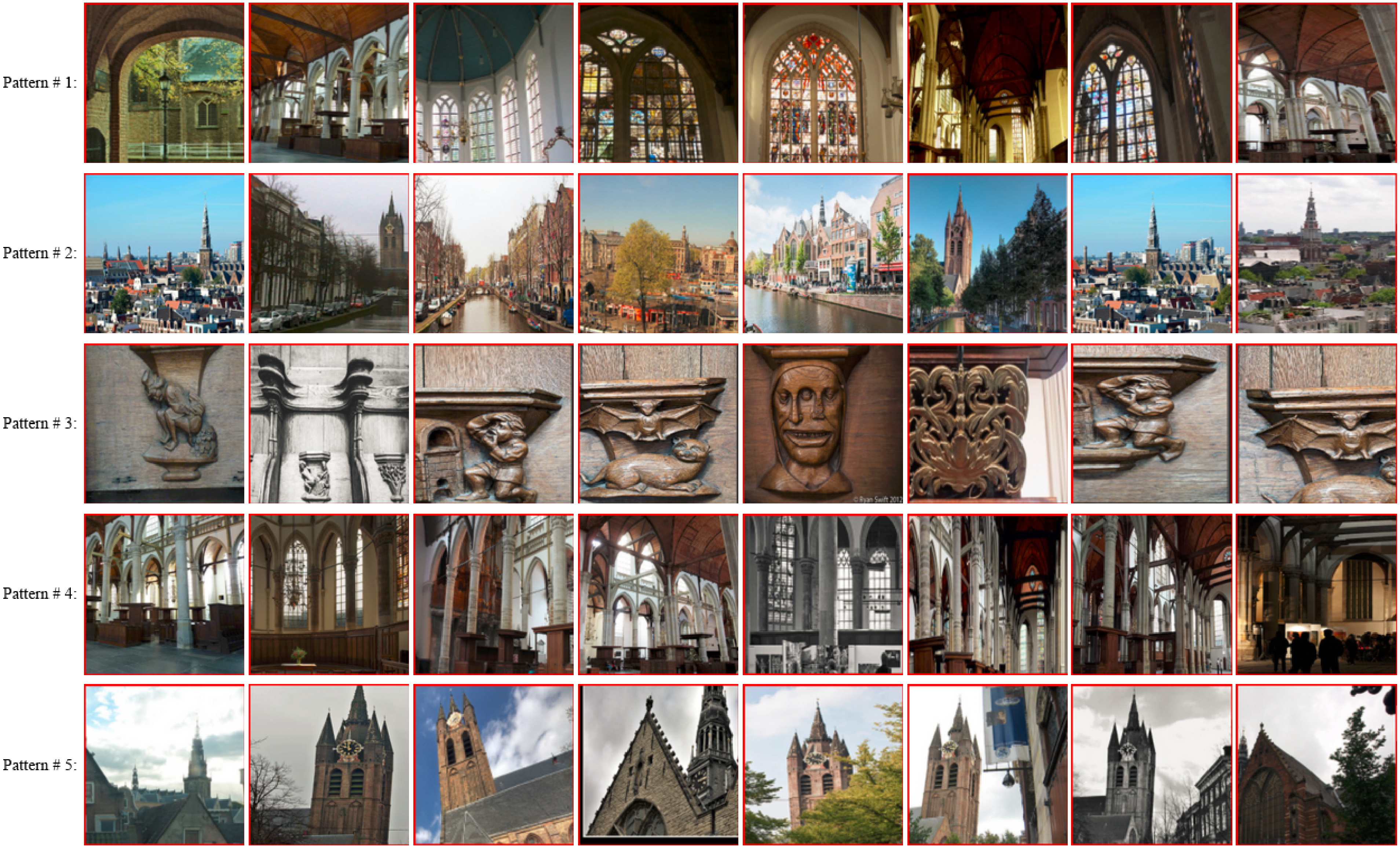


Research: Find unique visual patterns for each language



Li, Ellis, Chang, arXiv, 2016

Unique Visual Patterns for Dutch *Old Churches*



Unique Visual Patterns for Spanish ***Old Churches***

How do images of the same concept vary in different languages?

Ontology Structure

- flora
- flower
- fly
- fog
- food
  - bad_food
  - delicious_food
  - exotic_food
  - good_food
  - healthy_food
  - natural_food
  - sweet_food
  - traditional_food

MVSO / F / food / good_food

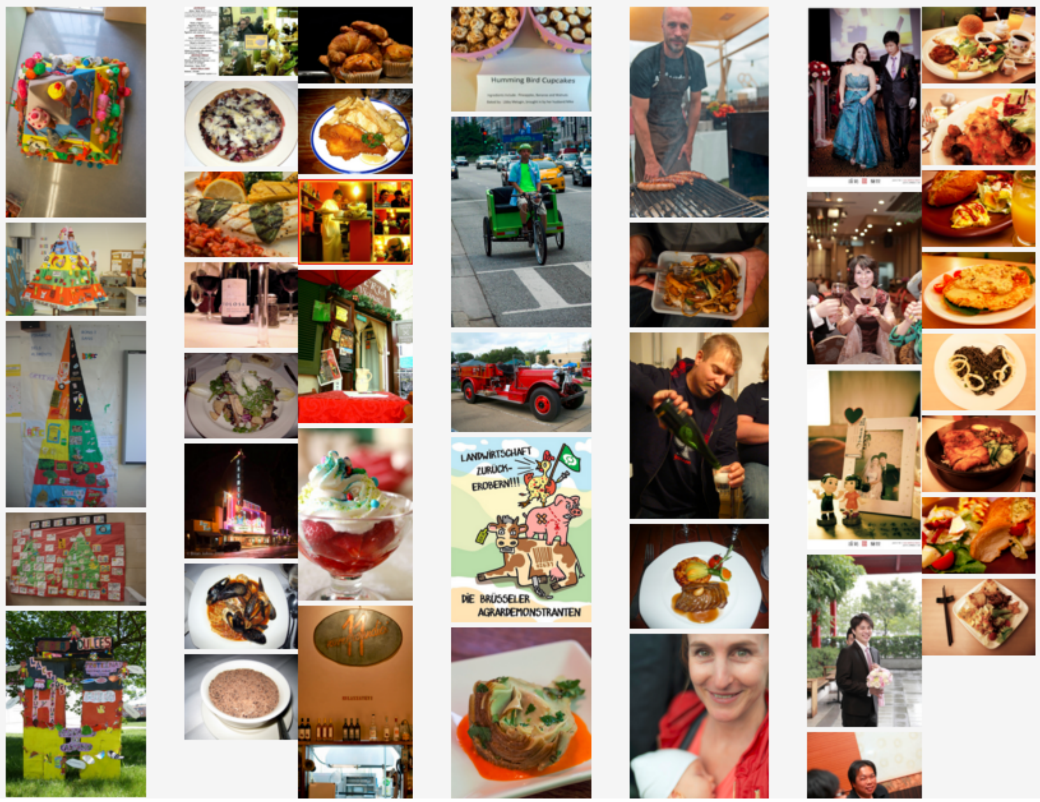
Children of Cluster

cibo_buono good_food
alimentos_buenos 好_料理
lekker_eten buona_tavola 好_食物

ANP Color Representation

- ENGLISH
- CHINESE
- DUTCH
- SPANISH
- ITALIAN

ES IT EN NL ZH



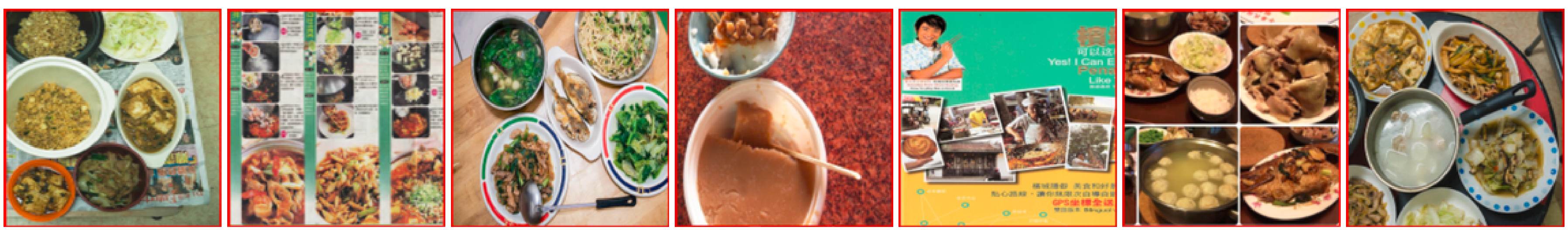
Unique Visual Patterns for Chinese ***Good Food***

Pattern # 1:

Pattern # 2:

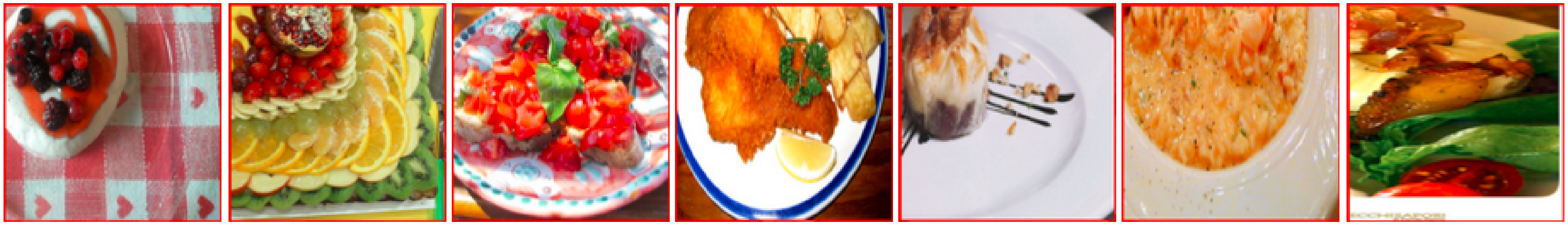
Pattern # 3:

Pattern # 4:



Unique Visual Patterns for Italian ***<u>Good Food</u>***

Pattern # 2:



Pattern # 3:

Pattern # 4:

Pattern # 5:

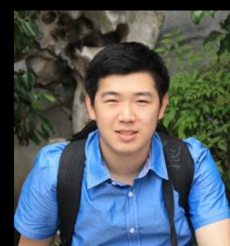
NEWS ROVER

New frontier of ***multi-source***, ***multi-modal***, ***personalized*** news exploration

Shih-Fu Chang

Joe Ellis

Brendan Jou

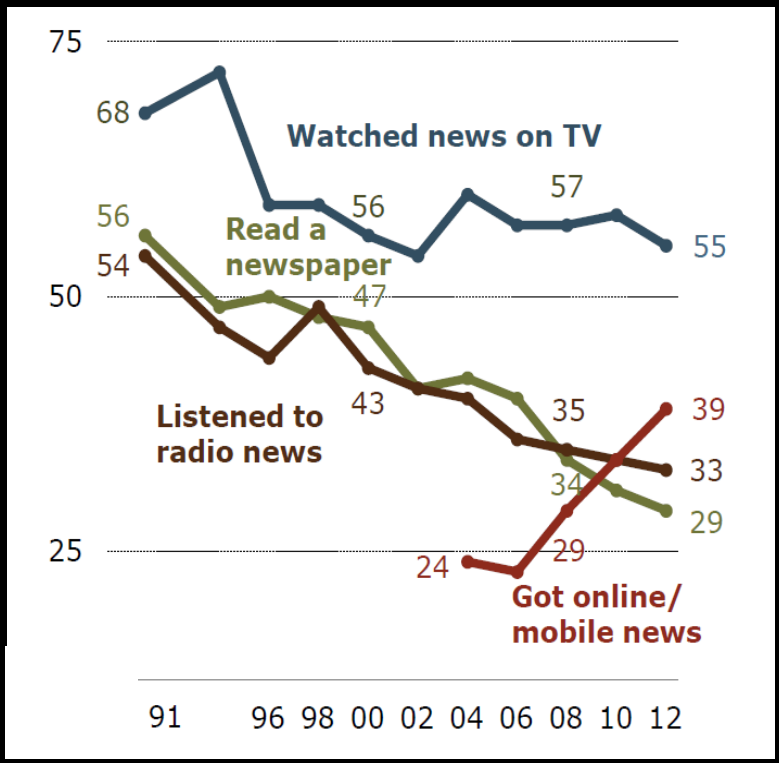


Hongzhi Li

Dan Morozoff-Abezgauz

User trend in news consumption

Where did you get your news yesterday?



Changing

- Towards mobile and social
- User interest guided by topic not by channel
- Mix any source (article, video, tweet)
- Access on any device, time, place

Source: PEW RESEARCH CENTER 2012 News Consumption Survey

NEWS ROVER

New frontier of *multi-source*, *multi-modal*, *personalized* news exploration

How will we watch news in the future?

NBA finals | tech | elections | elections | olympics | Greek riots

Topical + Personal

Organized by personal interest and trendy topics

Mix across multiple sources

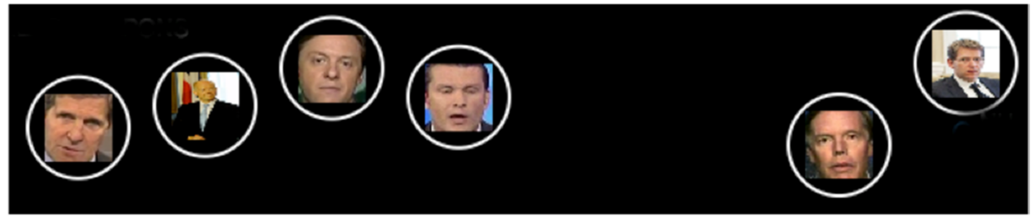
Anywhere anytime

Videos analyzed and processed in the cloud, then consumed anywhere

Navigating the News

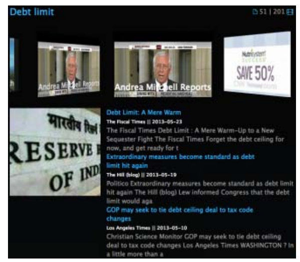
Who?

Name Extraction & Normalization

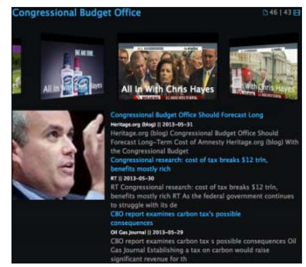


What?

Topical Organization



'Debt Limit'

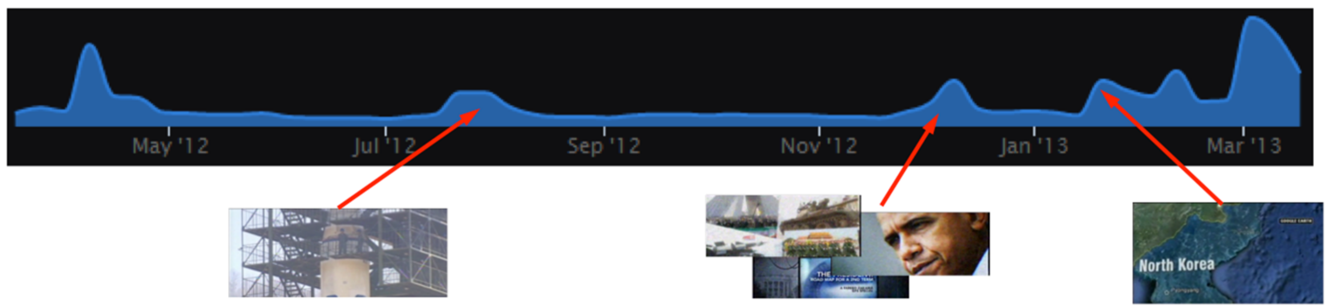


'Cong. Budget Office'

'Consumer Debt'

When?

Coverage Trendline



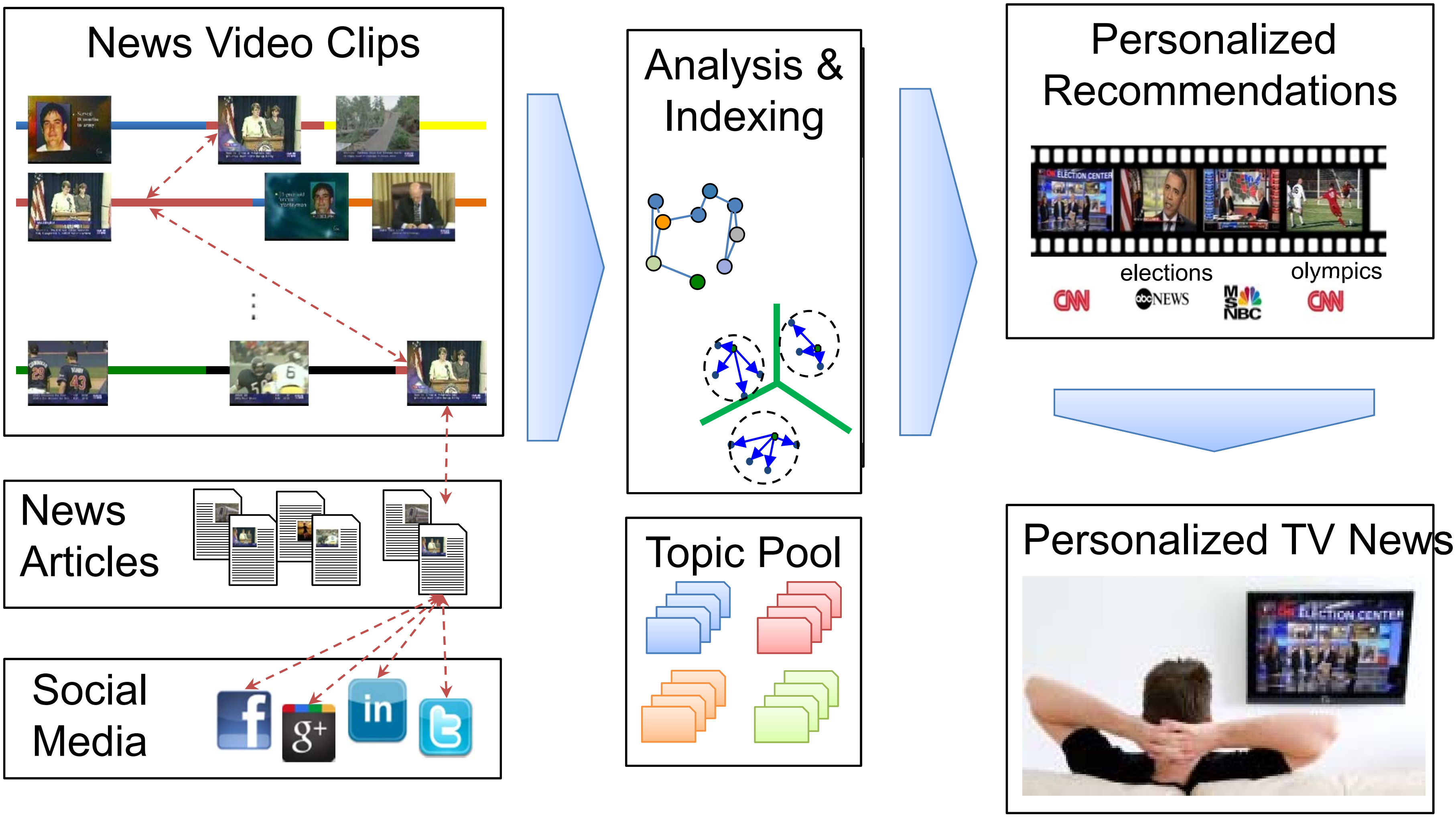
Where?

Event Geo-visualization

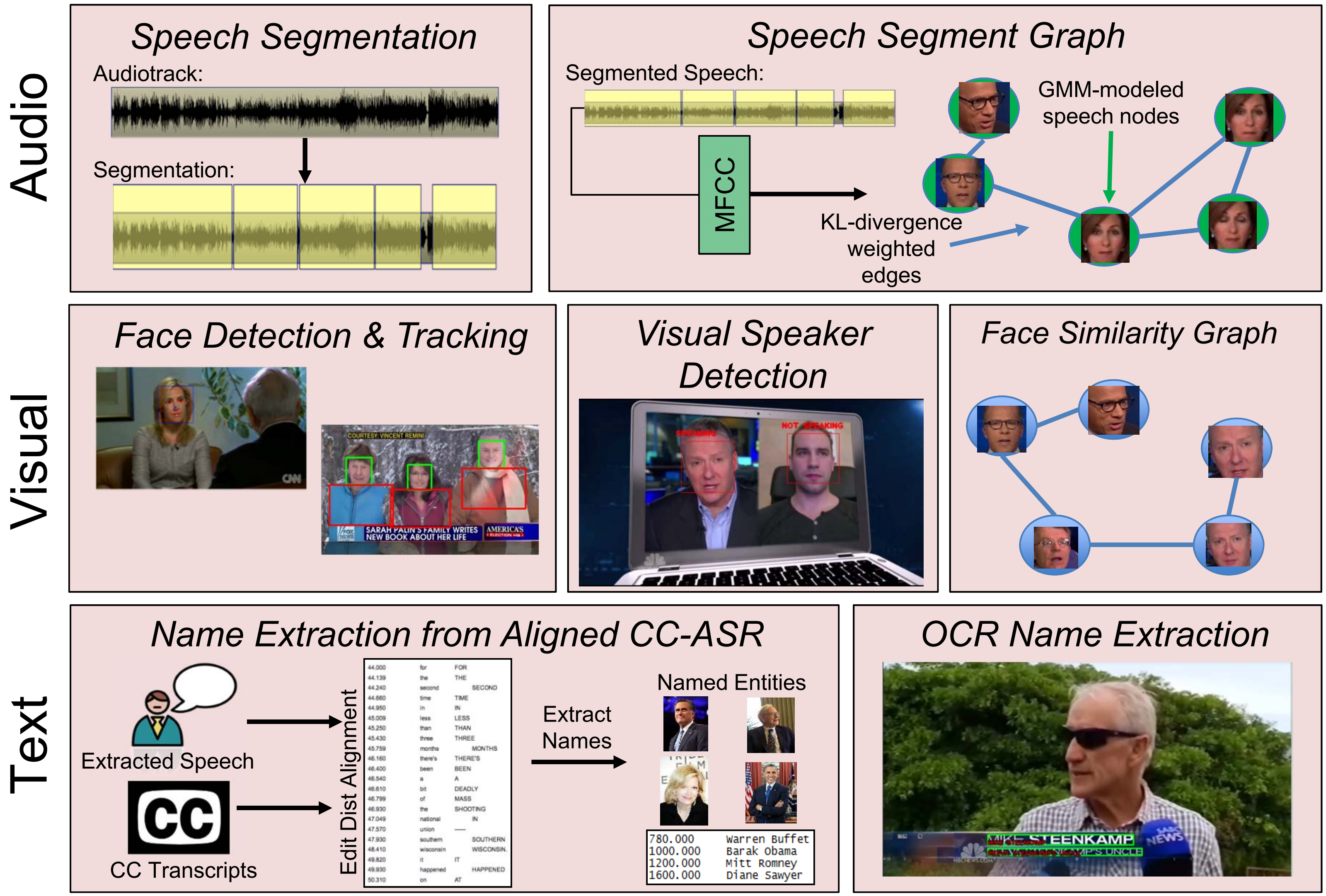
'Same-sex Marriage'

'France/Mali' conflict

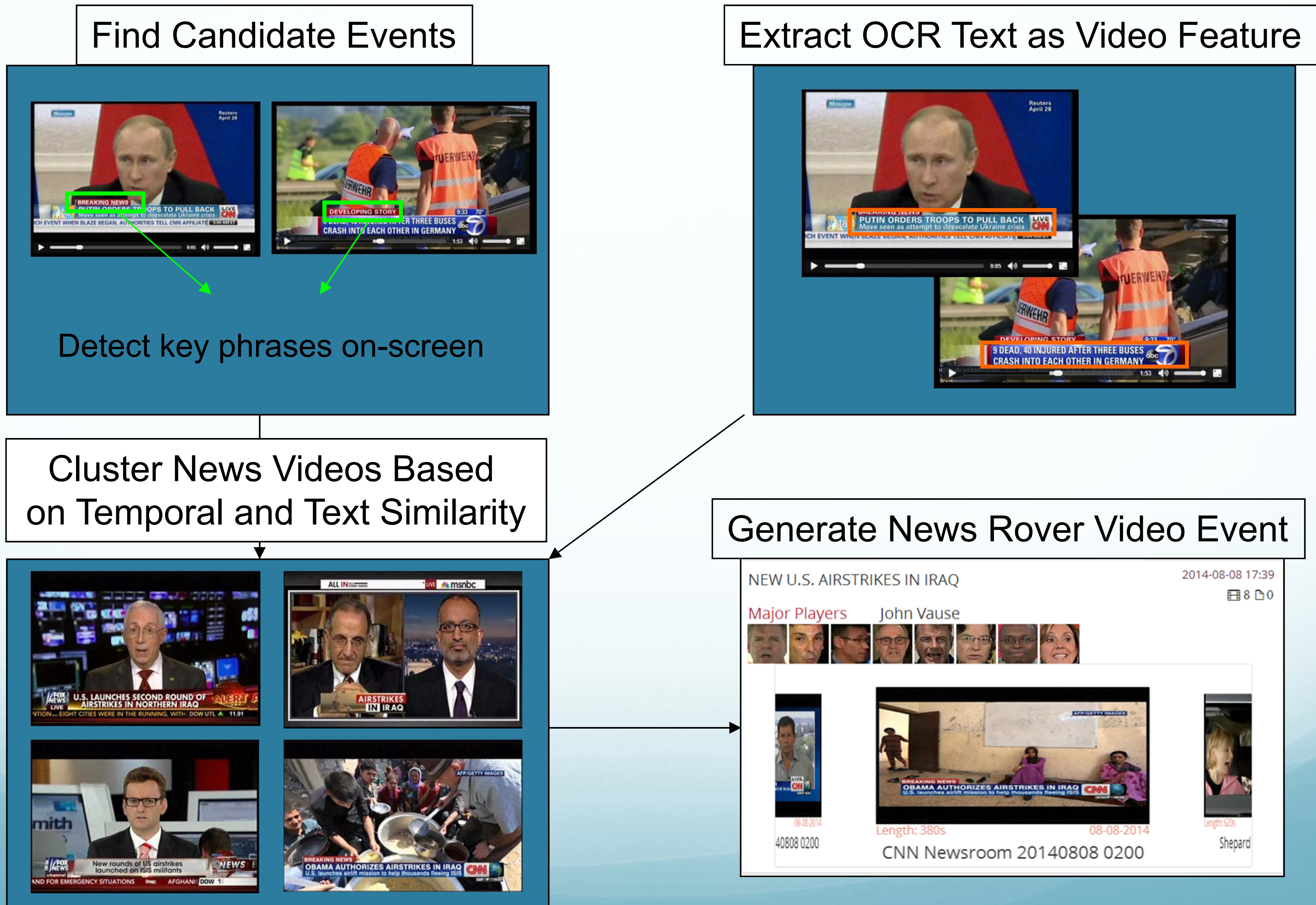
Multi-Source Aggregation



Research: Multimodal Analysis "Who Said What"



Video-Based News Events



NewsRover Live Recording System

100 TV channels recorded continuously

Linked to Online News and Twitter topics

- Per day:
  - **110** hours video recorded
  - **1,380** video stories indexed
  - **460** Google News topics crawled
  - **550** Twitter topics crawled
- Total size:
  - **28,000** hours video recorded
  - **464,000** video stories
  - **24,500** News topics
  - **80,000** Twitter topics