Exploring Image, Video and Multimedia in Large Data Applications

Prof. Shih-Fu Chang
http://www.ee.columbia.edu/dvmm
January 29th, 2016
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 ...

Photos Alone = 1.8B+ Uploaded & Shared Per Day...
Growth Remains Robust as New Real-Time Platforms Emerge

Daily Number of Photos Uploaded & Shared on Select Platforms, 2005 – 2014YTD

Source: KPCB estimates based on publicly disclosed company data, 2014 YTD data per latest as of 5/14
S.-F. Chang
Researchers Announce Advance in Image-Recognition Software

By JOHN MARKOFF  NOV. 17, 2014

How Facebook is teaching computers to see

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

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)

S.-F. Chang
ImageNet: Recognize > 1,000 object categories

Variety of object classes in ILSVRC

- **DET**
  - birds
    - bird
  - bottles
    - bottle
  - cars
    - car

- **CLS-LOC**
  - flamingo
  - cock
  - ruffed grouse
  - quail
  - partridge
  - pill bottle
  - beer bottle
  - wine bottle
  - water bottle
  - pop bottle
  - race car
  - wagon
  - minivan
  - jeep
  - cab 6
Rapid Progress in Recent Years:

Classification/Detection Task

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

Classification Error

<table>
<thead>
<tr>
<th>ILSVRC year</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification error</td>
<td>0.28</td>
<td>0.20</td>
<td>0.16</td>
<td>0.12</td>
<td>0.07</td>
<td>0.056</td>
</tr>
</tbody>
</table>

Localization Error

<table>
<thead>
<tr>
<th>ILSVRC year</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Localization error</td>
<td>0.43</td>
<td>0.34</td>
<td>0.30</td>
<td>0.25</td>
<td>0.09</td>
</tr>
</tbody>
</table>

1.9x decrease in classification error
2.8x decrease in localization error
But Use with Caution!

“Smart” Software Can Be Tricked into Seeing What Isn’t There

Nguyen, et al, CVPR 2015
Research Issue: Complexity

- 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.
Replace fully connected layer with structured component:

\[ h(x) = \phi(RDx) \]

\( \phi(\cdot) \): a nonlinear activation function

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

\[
R = \text{circ}(r) := \begin{bmatrix}
    r_0 & r_{d-1} & \cdots & r_2 & r_1 \\
    r_1 & r_0 & r_{d-1} & \cdots & r_2 \\
    \vdots & r_1 & r_0 & \cdots & \vdots \\
    r_{d-2} & \cdots & \cdots & r_{d-1} \\
    r_{d-1} & r_{d-2} & \cdots & r_1 & r_0
\end{bmatrix}
\]

- **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.

(and also works proposing structured efficient DL)
Circulant Neural Networks (cont’d)

<table>
<thead>
<tr>
<th>Method</th>
<th>Time</th>
<th>Space</th>
<th>Time (Learning)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional</td>
<td>$\mathcal{O}(d^2)$</td>
<td>$\mathcal{O}(d^2)$</td>
<td>$\mathcal{O}(ntd^2)$</td>
</tr>
<tr>
<td>Ours</td>
<td>$\mathcal{O}(d \log d)$</td>
<td>$\mathcal{O}(d)$</td>
<td>$\mathcal{O}(ntd \log d)$</td>
</tr>
</tbody>
</table>

1000x less parameters in fc

- 4K × 4K layer: 64MB → 16KB

<table>
<thead>
<tr>
<th>Method</th>
<th>Top-5 Err</th>
<th>Top-1 Err</th>
<th>Memory</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlexNet (rand)</td>
<td>33.5%</td>
<td>61.7%</td>
<td>233.2MB</td>
</tr>
<tr>
<td>C-AlexNet (rand)</td>
<td>35.2%</td>
<td>62.8%</td>
<td>20.5MB</td>
</tr>
<tr>
<td><strong>AlexNet</strong></td>
<td>17.1 %</td>
<td>42.8%</td>
<td>233.2MB</td>
</tr>
<tr>
<td><strong>C-AlexNet</strong></td>
<td>19.4 %</td>
<td>44.1%</td>
<td>20.5MB</td>
</tr>
<tr>
<td>More parameters</td>
<td>17.8 %</td>
<td>43.2%</td>
<td>20.7MB</td>
</tr>
<tr>
<td>Reduced AlexNet</td>
<td>37.2 %</td>
<td>65.3%</td>
<td>20.7MB</td>
</tr>
</tbody>
</table>
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

• Detecting complex events in ~ 100,000 videos

K.-T. Lai; F. X. Yu; M.-S. Chen; S.-F. Chang. CVPR 2014
What Salient Concepts Are Associated with Each Video Event?

- **Event**
  - Object
  - Scene
  - Action
  - Scene
  - Action

- **Birthday Party**
  - Object
  - Cake
  - Candle
  - Candy
  - Gift
  - Food

- **Activity**
  - Lighting Candles
  - Cutting Cake
  - Drinking
  - Dancing
  - Home
  - Restaurant
  - Park
Inspiration from Crowd – WikiHow

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

- Sports and Fitness
  - For the Fan
    - Fans of Baseball • Fans of Football • Fans of Hockey • Fans of Basketball
  - Individual Sports
    - Bicycling • Racket Sports • Surfing • Swimming and Diving • Parkour • Inline and Roller Skating • Trampoline • Cue Sports • Bowling • Combat Sports • Archery • Ice and Figure Skating • Snow Sking • Water Skiing • Track and Field • Snowboarding • Ballon Twirling • Golf • Skateboarding • Air Sports • Kayaking Canoeing and Rowing • Gymnastics
  - Outdoor Recreation
    - Backpacking and Hiking • Camping • Fishing • Hunting • Outdoor Safety • Climbing • Scouting & Other Youth Groups • Winter Outdoor Activities • Caving and Spelunking • Guns and Shooting • Enjoying the Great Outdoors • Orienteering and Outdoor Treasure Hunting • Knot Tying • Enjoying the Beach
  - Personal Fitness
    - Calisthenic Exercises • Cardio Exercises • Warm Ups Stretching and Flexibility • Gym • Hooping • Running for Fitness • Walking for Fitness • Motivation to Exercise • Building Muscle & Strength • Pilates

- Bicycling
  - BMX
    - Learning to Bicycle Ride
    - Mountain Biking
    - Unicycling

How to Assemble a BMX Bike

Building a bike can fun and easy but if you don’t know how to do it can be frustrating and hard. This guide promises that you will have a good experience.
Columbia Video EventNet Ontology (500 events, 4,800 concepts)

Demo
Demo2

Ye, et al, EventNet, ACM MM ’15, eventnet.ee.columbia.edu
Video in Large Data System: Traffic Cams
Urban Sensing: City webcam and social media

- 400+ traffic webcams in NYC: [http://nyctmc.org/](http://nyctmc.org/)
- Resolution 352x240. Framerate between 1fps and 1 frame every 3s.
- No ground truth. Event calendars from OpenData and DOT API
  - [https://developer.cityofnewyork.us/api/events-calendar](https://developer.cityofnewyork.us/api/events-calendar)

<table>
<thead>
<tr>
<th>event</th>
<th>date</th>
<th>time</th>
<th>location</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBGB Music Festival</td>
<td>12 Oct</td>
<td>10am-7pm</td>
<td>Broadway 51 Street</td>
</tr>
<tr>
<td>Hispanic Parade</td>
<td>12 Oct</td>
<td>12pm-5pm</td>
<td>5th Avenue</td>
</tr>
<tr>
<td>Columbus Day Parade</td>
<td>13 Oct</td>
<td>11am-5pm</td>
<td>5th Avenue</td>
</tr>
<tr>
<td>Saint Patrick's Day Parade</td>
<td>17 Mar</td>
<td>12pm-5pm</td>
<td>5th Avenue</td>
</tr>
<tr>
<td>Million March NYC Protest</td>
<td>13 Dec</td>
<td>2pm-5pm</td>
<td>Washington Square Park, 5th Avenue, Foley Square</td>
</tr>
</tbody>
</table>
Example Tweets (Wu & Kankanhalli, WWW’15)

2011  2015-10-31 05:45:21 Binnen... Incl honered guest bandje ? 1 day 2 #tcsnycmarathon @ Times Square, New York City https://t.co/ew4buSUhDR  40.758895 -73.985131

2074  2015-10-31 06:07:34 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  2015-10-31 06:55:00 Not a bad view from my hotel room watching the sun come up!! Off to the marathon expo to collect my¡ https://t.co/HWr90OH88Ht  40.75807569 -73.9753008
Social Crowd Reporting (e.g., Waze)
Computer Vision Tools

**car detection - success in most cases:**

![Example Images]

**car detection - some missing detections:**

![Example Images]
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.
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
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.
Beyond Semantics

Aesthetics

Interestingness

Beyond Semantics

Emotion

Style

Others: Creativity, Intent, Memorable ...

(CVPR 2014 Tutorial)
Culture plays an important role in visual presentation.
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

Build Sentiment Ontology

Select Concepts

Analyze tags with strong sentiments

Borth, Ji, Chen, Breuel, Chang, *Large-Scale Visual Sentiment Ontology*, ACM Multimedia 2013
Frequent Photo Tags Related to Emotions
Currently, we have found 3000+ ANPs.

Discovered Sentiment-Adjective-Noun Pairs in Social Media Photos
Visual Sentiment Ontology (Browser)

Ontology :: Emotional Mapping

Selected Emotion:

<table>
<thead>
<tr>
<th>ANPs</th>
<th>sentime</th>
<th>emotion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. happy smile</td>
<td>1.92</td>
<td>0.388</td>
</tr>
<tr>
<td>2. innocent smile</td>
<td>1.92</td>
<td>0.376</td>
</tr>
<tr>
<td>3. happy christmas</td>
<td>2.0</td>
<td>0.373</td>
</tr>
<tr>
<td>4. happy father</td>
<td>2.0</td>
<td>0.358</td>
</tr>
<tr>
<td>5. happy wedding</td>
<td>1.72</td>
<td>0.348</td>
</tr>
<tr>
<td>6. friendly smile</td>
<td>1.92</td>
<td>0.346</td>
</tr>
<tr>
<td>7. delicious cupcake</td>
<td>1.71</td>
<td>0.341</td>
</tr>
<tr>
<td>8. shy smile</td>
<td>0.62</td>
<td>0.340</td>
</tr>
<tr>
<td>9. charming smile</td>
<td>1.92</td>
<td>0.339</td>
</tr>
<tr>
<td>10. happy birthday</td>
<td>1.79</td>
<td>0.337</td>
</tr>
<tr>
<td>11. warm smile</td>
<td>1.92</td>
<td>0.336</td>
</tr>
<tr>
<td>12. happy mother</td>
<td>2.0</td>
<td>0.333</td>
</tr>
<tr>
<td>13. happy halloween</td>
<td>1.81</td>
<td>0.331</td>
</tr>
<tr>
<td>14. delicious drink</td>
<td>1.59</td>
<td>0.330</td>
</tr>
<tr>
<td>15. happy heart</td>
<td>2.0</td>
<td>0.329</td>
</tr>
<tr>
<td>16. happy kids</td>
<td>2.0</td>
<td>0.323</td>
</tr>
<tr>
<td>17. healthy food</td>
<td>1.69</td>
<td>0.312</td>
</tr>
<tr>
<td>18. happy guy</td>
<td>1.61</td>
<td>0.312</td>
</tr>
<tr>
<td>19. fresh food</td>
<td>1.59</td>
<td>0.307</td>
</tr>
<tr>
<td>20. delicious pie</td>
<td>1.76</td>
<td>0.296</td>
</tr>
</tbody>
</table>
Visual Sentiment Ontology (Browser)
**THE MULTILINGUAL VISUAL SENTIMENT ONTOLOGY**

**MVSO Pipeline**

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

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, 欢乐, 无聊)

(v) German (top: Angst, Freude, Interesse)

(vi) Arabic (top: الحزن, الحب, 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’

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

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

![Images of concepts](image1.png)
How do images of the same concept vary in different languages?
“Old Church” Images: Cultural Influences?

Dutch

German

Italian

Spanish
Research: Find unique visual patterns for each language

1, 0, 0, 1, 0, 1, 0, ... 0, 0, 0
Each patch encoded as binary transaction item set

Deep neural network → Filter response maps → NonMax-Suppression over each response map

256-dimensional feature vector
Concatenate filter responses

Li, Ellis, Chang, arXiv, 2016
Unique Visual Patterns for Dutch *Old Churches*

Pattern # 1:

Pattern # 2:

Pattern # 3:

Pattern # 4:

Pattern # 5:

S.-F. Chang
Unique Visual Patterns for Spanish *Old Churches*
How do images of the same concept vary in different languages?

S.-F. Chang
Unique Visual Patterns for Chinese *Good Food*
Unique Visual Patterns for Italian *Good Food*

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-Abezgaus
User trend in news consumption

Where did you get your news yesterday?

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Watched news on TV</td>
<td>68</td>
<td>65</td>
<td>56</td>
<td>56</td>
<td>57</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read a newspaper</td>
<td>54</td>
<td>56</td>
<td>47</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listened to radio news</td>
<td>56</td>
<td>54</td>
<td>43</td>
<td>35</td>
<td>35</td>
<td>39</td>
<td>33</td>
<td>39</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>Got online/mobile news</td>
<td>25</td>
<td>29</td>
<td>24</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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
How will we watch news in the future?

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

When?
Coverage Trendline

Where?
Event Geo-visualization

When?
‘Debt Limit’
‘Cong. Budget Office’
‘Consumer Debt’

What?
‘Same-sex Marriage’
‘France/Mali’ conflict

Where?

Who?

When?

What?

Where?
Multi-Source Aggregation

News Video Clips

News Articles

Social Media

Analysis & Indexing

Topic Pool

Personalized Recommendations

Personalized TV News
Research: Multimodal Analysis “Who Said What”

Audio

- **Speech Segmentation**
  - Audiotrack:
  - Segmentation:

- **Speech Segment Graph**
  - Segmented Speech:
  - MFCC
  - KL-divergence weighted edges
  - GMM-modeled speech nodes

Visual

- **Face Detection & Tracking**
  - Face Detection & Tracking
  - Visual Speaker Detection
  - Face Similarity Graph

Text

- **Name Extraction from Aligned CC-ASR**
  - Extracted Speech
  - CC Transcripts
  - Edit Dist Alignment

- **OCR Name Extraction**
  - Extract Names
  - Named Entities
  - Warren Buffett
  - Barack Obama
  - Diane Sawyer
Video-Based News Events

Find Candidate Events
- Detect key phrases on-screen

Extract OCR Text as Video Feature

Cluster News Videos Based on Temporal and Text Similarity

Generate News Rover Video Event
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