Computer Vision Review and Outlook
What is computer vision?

– “Making computers see”
Computer Vision

• Vision is easy, right? Just open your eyes!
  – No, it’s a hard problem…
  – Much of your very complex brain is devoted to doing vision
  – It involves cognition, navigation, manipulation and learning
    • Not just simple “match a feature vector to a database” tasks

• CV is about interpreting the content of images (incl. video)
  – Field is ~50 years old
  – Originally a child of AI
  – Now closely related to several other research fields
Closely related fields of research

- Image processing
- Human vision
- Computer graphics
- Machine learning
- Computer vision
- Pattern recognition
- AI
- Statistics
- Robotics
- Optics
Broad participation

• Computer vision research takes place in several academic departments
  – Computer Science
  – Electrical Engineering
  – Computer Engineering
  – Statistics
  – Psychology
  – Applied Mathematics

• It is inherently a multidisciplinary field
  – Computer vision education is a challenge
    • It’s impossible to know everything!
What does Computer Vision do?

- 3D models of objects
- Object recognition
- Navigation
- Event/action recognition
...
Possible outputs

• Objects
  – Cat, chair, window, star, bush, water, a shoe, my mother…

• Properties
  – Big, bright, yellow, fast, graspable, moving…

• Relations
  – In front, behind, on top, next to, larger, closer, identical…

• Shapes
  – Round, rectangular, star-shaped, symmetric…

• Textures
  – Rough, smooth, irregular…

• Movement
  – Turning, looming, rolling…
“J548043”
“A harbor with many dozens of boats; water is calm and glassy; masts are all vertical; mountains in background, blue sky with a touch of clouds…”

or

“Boat”

or

“Outdoors”
“Hallway straight ahead”
“Angry”

“Surprised”

“Happy”

“Disgusted”
Why is vision difficult?

Consider the input...

Not this
But this...
Some possible outputs

- Depth or segmentation
- Object pose (facing away, facing forward)
- Object recognition
- Action understanding

Oulu 6.2012
What your brain does

Almost certain to be Bill Clinton
Gray hair
Dark circular overlay
Right ear
Neck
White shirt
Armani suit
Person contour
CNN caption (Washington 1995?)
Clinton occluding Monica

Clinton greeting Lewinsky

Person with glasses in crowd
Pony tail
Dark brown hair
Right eye (open)
Left eye (open)
Nose
Cheek
Monica’s mouth (smiling)
Lapel
Illuminated from above
Necklace
Woman’s dress suit

Monica Lewinsky
Variation in appearance
What makes computer vision hard?

- Underspecified problem!
  - 3D world projected onto 2D sensor(s)
- Environment
  - Lighting, background, movement, camera
- Varying appearance of objects
- Calibration, FOV, camera control, image quality
- Computational complexity (speed of processing)
- Etc., etc….

*Robustness is the primary challenge to computer vision!*
CV Research: The Past

• Main focus of research
  – Camera (sensor) calibration and image formation
  – Binary image analysis
  – Edge detection and low-level filtering
  – Color representation and segmentation
  – Texture description and segmentation
  – Depth/shape from X
  – Stereopsis
  – Optical flow
  – Motion computation
  – Object matching, detection, and recognition
  – 3D sensing and shape description
  – Object and scene tracking

High-level vision
(cognition, semantics)?
CV Research: Past and Present

- Main driving applications:
  - Biological vision
  - Robotics (factory, mobile)
  - Remote imaging
  - Medical imaging
  - Security and surveillance
  - Multimedia database query (digital libraries)
  - Computer graphics, special effects
  - Multimedia database query (digital libraries)
  - HCI
  - Consumer applications
Progress in Computer Vision

• First generation: Military/Early Research
  – Few systems, each custom-built, cost $M$s
  – “Users” have PhDs
  – 1 hour per frame

• Second generation: Industrial/Medical
  – Numerous systems, 1-1000 of each, cost $10K$s
  – Users have college degree
  – RT with special hardware

• Third generation: Consumer
  – 100000(00) systems, cost $100$s
  – Users have little or no training
  – RT in software
Examples

CMU NavLab

Cognex

Microsoft Kinect
Computer Vision Today

• Lots of progress in low- to mid-level vision
  – Calibration
  – Computing features: edges, depth, motion, color, etc.
  – Shape and texture representations
  – Tracking, motion
  – Basic segmentation, classification
  – Simple recognition

• Trends
  – More thorough experimentation, comparisons (e.g., FRVT)
  – Probabilistic models, data intensive, learning
  – Variational and level-set methods
  – Real-time systems

CVPR demo sessions are finally interesting!
Where is the field heading?

• Wider variety of driving applications
  – Entertainment
  – Communication
  – Consumer electronics (images everywhere)

• Continued progress toward robust low/mid-level vision
  – Tracking, depth, etc.

• Multiple sensing modes
  – Integration of vision, speech, language

• Connect back to AI
  – Need to involve semantics and context again
    • Completely bottom-up data-driven methods aren’t sufficient

Mobile!
Mobile computer vision is different

• Traditional CV scenario
  – One or a few expensive cameras
  – Connected to a dedicated workstation(s)
  – Fixed location (except in robotics)
  – Isolated from other devices
  – Vision-only (except in robotics)
  – Few or no power constraints
  – Driven by military or industrial applications

• Mobile computing is changing all this, and is now driving the field of computer vision (research and practice) in new directions
Mobile computer vision

• Mobile CV scenario
  – One or dozens or thousands of inexpensive cameras
  – Embedded in everyday device
  – Mobile, accessible
  – Connected
  – Associated sensor suite
Mobile CV – driving applications

• Traditional CV application areas:
  – Military
  – Manufacturing
  – Robotics
  – Medical
  – Surveillance

• Mobile vision driving application areas:
  – **Photography** (people, pets, landscapes, tourism, etc.)
  – Visual search
  – Augmented reality
  – Personal navigation
  – Life recording and annotation
  – Collaboration
Mobile computer vision

• Mobile computer vision is not just “making vision routines fast/small enough to run on mobile devices”
  – Although this is part of the engineering challenge

• Ideally, mobile CV should leverage
  – Mobility
  – Convenience, accessibility of camera
  – Connectedness (broadband access)
  – Integrated sensors
  – Consumer-oriented opportunities (personal and social)
  – Collaboration
Mobile CV—opportunities and challenges

• Huge opportunities, due to
  – Ubiquity of devices, size/portability/availability/affordability
  – Embedded sensors:
    • Camera(s), accelerometer, magnetometer, gyroscope, GPS, …
  – Connectivity, potential for collaboration
  – Consumer-driven interests and applications

• But limited by
  – Hardware capabilities (processing speed, memory, etc.)
  – Control and access by applications
  – Camera quality
    • Optics, sensor array, rolling shutter, video compression, …
  – Bandwidth
  – Power!!
Challenges for mobile computer vision

• Mobile computer vision cannot be just “making CV algorithms and applications run on less powerful devices”
  – Must embrace the differences and opportunities of this new paradigm

• Fully leverage the multi-sensor capabilities of the device
  – Two (or more) cameras, inertial sensors, magnetometer, GPS, etc.

• Embrace the cloud as a integral component in CV

• Realize the potential of collaboration – leverage the ubiquity and connectedness of devices
  – E.g., ad hoc networks of mobile CV devices

• Build software infrastructure to support the mobile CV paradigm
  – Beyond OpenCV, AR SDKs, etc.
Collaborative mobile applications

• Mobile devices have great collaboration potential, beyond social networking
• Mobile collaborative sensing = sensor networks comprising people with their smartphones

E.g.:
• Collaborative driver assistance
  – E.g., traffic advisory, monitoring road conditions, parking space availability, taxi availability, …
• Collaborative presence
  – E.g., richly experience an event (concert, demonstration, lecture), live or delayed, using mobile phone video and audio of the people at the event
Some recent key indicators

- Google acquires CMU spinoff Pittpatt (2011)
  - Face recognition: Face Unlock in Android 4.0
- Lytro camera (2012) – “focus later”
  - From 2005 Stanford PhD dissertation
- Nokia acquires Swedish company Scalado (2012)
  - Mobile imaging: capturing, viewing, editing, sharing images
- Nokia PureView (2012)
  - 41 MP camera
- Facebook acquires Instagram for $1B (2012)
  - Consumer: imaging + social
- Google Glass (2013?)
  - A new consumer imaging opportunity?
Things to encourage

• Continue current trends (learning, prob. models, data)
  – Big lack of common data sets w/ground truth
• But avoid “blind application of X” (NNs, HMMs, SVMs, PR techniques)
  – Data is not enough!
• Competitions, data sets
  – Common task, common data (e.g., DARPA with speech rec.)
  – International, including industry
• Key application areas
  – Medical imaging, security and surveillance, digital libraries, robotics, HCI, art and entertainment, consumer/communication apps, health, education
Things to encourage (cont.)

• Focus on robustness
  – Need better definitions, standard ways to evaluate robustness
• Applications that keep a human in the loop
  – Augment, not necessarily replace

Key funding categories:

1. Core vision problems (stereo, motion, segmentation, etc.)
2. Fundamental methods (linear/nonlinear dynamic models, probabilistic models, Monte Carlo techniques, etc.)
3. Key application areas that drive #1 and #2