Semantically Linking Instructional Content

by Yekaterina Kharitonova

ISTA 352 / University of Arizona / October 5, 2012

Let's imagine...

Main Goal

To split-up a video into semantically meaningful clips based on the slide usage.

Main Task

Align a video to the slides used during the presentation.

Video: a sequence of video frames.

Slides: a JPG image for each slide.

What is a slide?

What is a slide?

1. **Feckless**: generally incompetent and ineffectual
2. **Foible**: a behavioral attribute that is distinctive and peculiar to an individual
3. **Fulmination**: thunderous verbal attack
4. **Gaucherie**: a socially awkward or tactless act
5. **Homiletics**: the act of preaching
6. **Imbroglio**: an intricate and confusing interpersonal or political situation

---

What is a slide?

1. **Feckless**: generally incompetent and ineffectual
2. **Foible**: a behavioral attribute that is distinctive and peculiar to an individual
3. **Fulmination**: thunderous verbal attack
4. **Gaucherie**: a socially awkward or tactless act
5. **Homiletics**: the act of preaching
6. **Imbroglio**: an intricate and confusing interpersonal or political situation


What is a slide?

Accolade: a tangible symbol signifying approval or distinction

Ambrosial: extremely pleasing to the taste

Canvass: to get the opinions (of people) by asking specific questions

Churl: a selfish person who is unwilling to give or spend

Cloture: a rule for limiting or ending debate in a deliberative body

Coda: the closing section of a musical composition

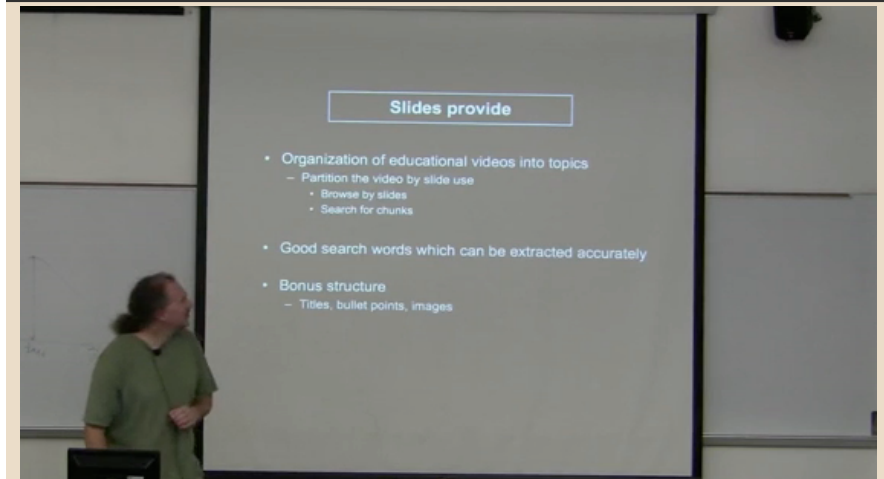**Countervail:** compensate for or counterbalance

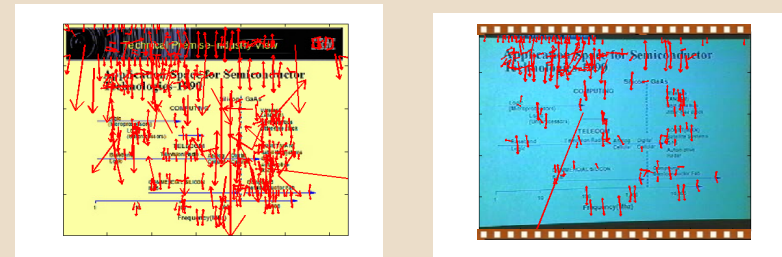Need a representation of a slide, which can be "taught" to a computer?

Where is the slide?
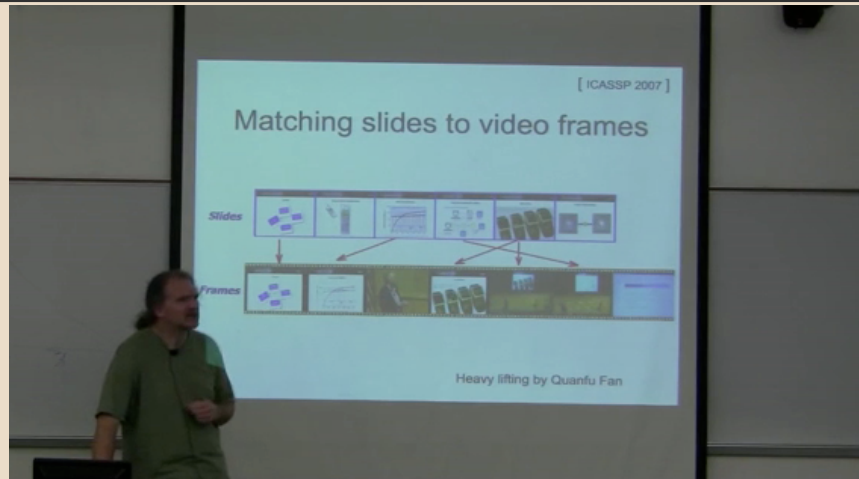
Where is the slide?

SIFT keypoints

- **S**cale-**I**nvariant **F**eature **T**ransformation
- Image keypoints – "interesting" points
  - o (x,y) position on the image
  - o Scale and Orientation
  - o 128-dimensional descriptor (texture around the keypoint)
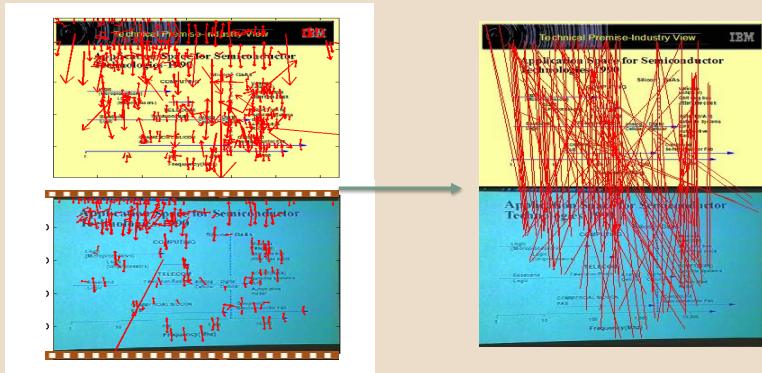
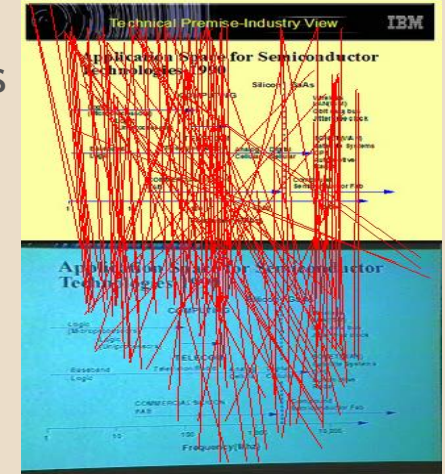# SIFT (examples)

Images courtesy Kobus Barnard and Quanfu Fan

Nearest Neighbor Match

- Match keypoints based on the descriptor

Nearest Neighbor Match

Can all these keypoint matches be correct?

Mapping the keypoints

- Mappings of points on a plane in 3D satisfy a simple relation (linear equation in homogeneous coordinates):

$$\lambda' \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Frame keypoint location (x',y') Homography, H Slide keypoint location (x,y)

Mapping the keypoints

- In homogenous coordinates, the slide-to-frame mapping is $\mathbf{X'} = H\,\mathbf{X}$

- The homography matrix, H, is a 3x3 matrix with 8 degrees of freedom

- Frame-to-slide mapping is similar: $\mathbf{X} = H'\,\mathbf{X'}$, where H' is the inverse slide-to-frame homography

Estimating the homography

- RANSAC: **RAN**dom **SA**mple **C**onsensus
- An iterative algorithm that estimates parameters of a model from a set of observed data (which contains outliers).

Estimating the homography

- RANSAC: **RAN**dom **SA**mple **C**onsensus
- An iterative algorithm that estimates parameters of a model from a set of observed data (which contains outliers).

Model: homography transformation
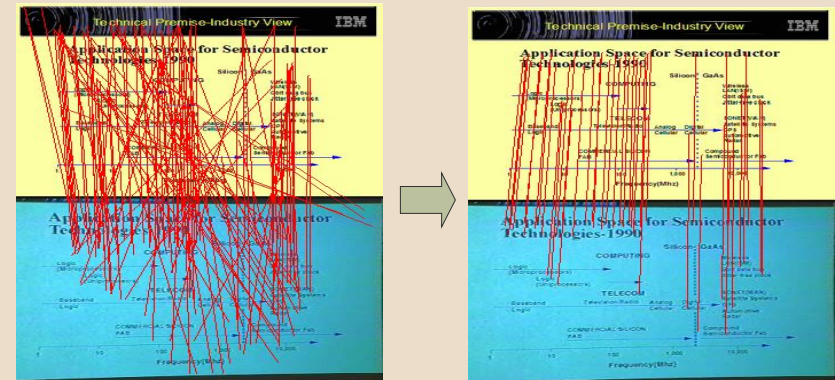
Model parameters: H matrix values

Observed data: matched keypoint locations

RANSAC

```
Loop
    - Select a set (Si) of random points
    - Fit a model (Hi) to Si
    If the model is good enough
        - Test all points against the model
        - Keep the best H
        - Re-estimate H using all points
```
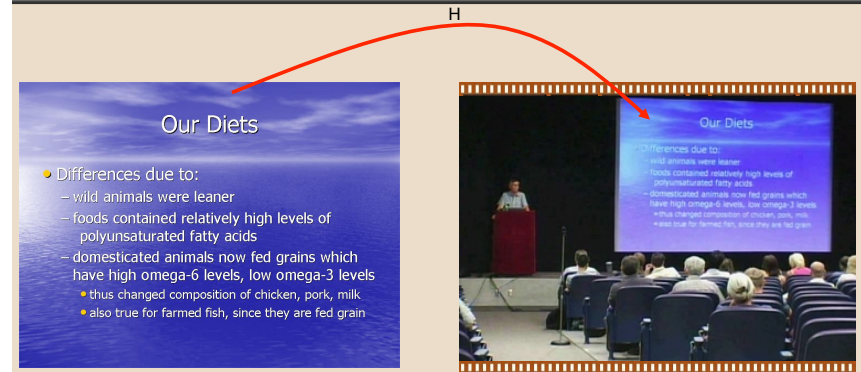
RANSAC

Without RANSAC
(Nearest neighbour)
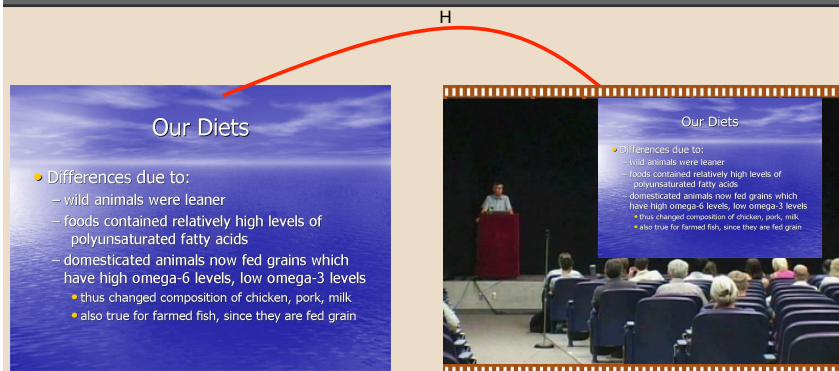
Matches limited to X'=HX
for some H

Homography Uses

- Backprojection

- Slide magnification

- Laser pointer-based bullet point magnification

Backprojection
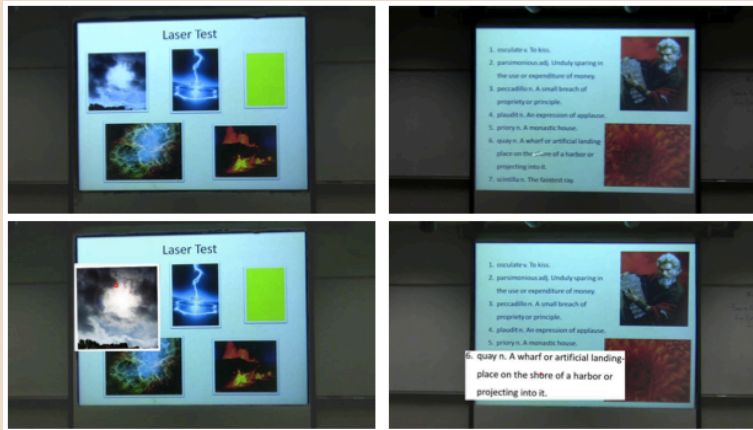

Backprojection

Slide Magnification

Images courtesy Andrew Winslow and Kobus Barnard

Bullet point magnification

Images courtesy Qiyam Tung and Alon Efrat

Challenges

- Animations (movies, cascading sequences of bullets)

- Blurry slides (out of focus, screen is brushed)

- Speaker occlusion

- Color shifts

Overview

- Motivation – quickly finding a relevant snippet of the video

- SIFT Keypoints

- Homography

- RANSAC

- Exploiting Homography

Contributors

Arnon Amir (IBM Almaden), **Kobus Barnard**, Troy Bowman, Joe Chitwood (KUAT), **Alexander Danehy, Benjamin Dicken, Alon Efrat**, Quanfu Fan, Sandiway Fong, Steven Gregory, **YK**, Derek Leverenz, Vivek Kumar, Ming Lin, Daniel Mathis, Adam McFarlin, Miguel Solano, Ranjini Swaminathan, Javad Taheri, Mohan Tanniru, Michael Thompson, Juhani Torkkola, **Qiyam Tung**, Gabriel Wilson, Andrew Winslow, Steve Zhou

Questions?