Survey Report

6.869 Advances in Computer Vision

Important Dates

Thursday, October 16	Survey report assigned
Thursday, October 30	Survey written report due in class and on stellar
Friday, October 31	Survey peer reviews assigned
Friday, November 7	Survey peer reviews due online
Tuesday, December 9	Final project report and presentation due in class

1 Survey report

We have covered a variety of topics in this course (with more to come). In the survey report, we want you to pick a topic that you find interesting, and become an expert in that area. Our hope is that your survey will give you the background knowledge for your final course project.

We have provided a list of topics below. Each topic suggests an implementation project and a paper list that you should read, but you should look for more related papers to read as well. Look through the topics, and pick one.

After reading the papers for your topic, we want you to write a two-page extended abstract of your topic. Your extended abstract should specifically introduce your topic, motivate why your topic is important, and explain how some of the methods in the literature approach this problem. Be sure to cite the papers you consulted. For an example, look at the first few pages of the paper "Deriving intrinsic images from image sequences" by Yair Weiss (online at `http://www.ai.mit.edu/courses/6.899/papers/13_02.PDF`), which does a good job at explaining intrinsic images.

Ultimately, your choice of survey topic will become the topic for your final project. In your report, you should also comment on how you might design your project, such as improvements to the method, changes to your dataset, or differences in model. We won't hold you to your ideas for the final project, but we are looking for evidence that you have critically thought about the project.

In your report, you should assume your reader has taken an advanced course in computer vision, but is not an expert in your chosen topic (yet!). When you turn in your paper, you should hand in a hard copy in class *with your name*, as well as upload an anonymous copy to stellar. We will distribute your anonymous report to two of your peers who will read and provide comments.

You can write your report independently, or you can work with *one* other person. The maximum group size is 2 people. A group only needs to turn in one report.

1.1 Report requirements

- Typed, two-page summary of your topic.
  - Single-spaced with a reasonable font size

- Double-column CVPR format
- One or two figures is okay, but not required
- Download style templates for LaTeX and Word at the CVPR website: `http://www.pamitc.org/cvpr14/author_guidelines.php`

- Report should:
  - introduce your topic
  - review methods to solve it
  - comment on your ideas for the final project.
- Maximum group size of 2 people
- You must submit two copies:
  - a hard copy in class *with your name*
  - a digital copy on Stellar *without your name*

2 Survey reviews

After you turn in your report, you will be assigned two reports from your peers that you should read and provide comments. For each report, you should write a short paragraph that discusses what the report does a good job at and where you think the report could be improved. You should specifically comment on how well the report introduces the topic and whether it motivates the problem as well how well it explains the methods to approach the problems.

3 Grading Rubric

The grade of your survey report will be partly based off the comments from your peers. We will assign points as following:

- 40% introduction of your topic
- 40% review of methods
- 10% proposals for final project
- 10% quality of reviews you write for other students

4 Topics

Below are the candidate projects for your survey and final project. Pick one that interests you. If you want to propose your own topic, please come talk to us during office hours, but we will only approve strong projects.

4.1 Motion magnification

Implement Eulerian motion magnification. Test it on many input videos (real and synthetic) using different parameters.

- Eulerian video magnification for revealing subtle changes in the world. Wu, Hao-Yu and Rubinstein, Michael and Shih, Eugene and Guttag, John V and Durand, Fredo and Freeman, William T

Related papers:

- Motion magnification. C Liu, A Torralba, WT Freeman, F Durand.
- Phase-Based Video Motion Processing. Neal Wadhwa, Michael Rubinstein, Frdo Durand, William T. Freeman
- Refraction Wiggles for Measuring Fluid Depth and Velocity from Video. T. Xue, M. Rubinstein, N. Wadhwa, A. Levin, F. Durand, W. T. Freeman
- The Visual Microphone: Passive Recovery of Sound from Video. Abe Davis, Michael Rubinstein, Neal Wadhwa, Gautham Mysore, Fredo Durand and William T. Freeman.

4.2 Object detection

Implement a ConvNet and test it on the MNIST digit recognition dataset or classify objects in a subset of ImageNet using pre-trained ConvNet features from Caffe.

- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

Related papers:

- R. Girshick, J. Donahue, T. Darrell, J. Malik. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation
- K. Simonyan, A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition
- Honglak Lee, Roger Grosse, Rajesh Ranganath, Andrew Y Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations
- Pedro F Felzenszwalb, Ross B Girshick, David McAllester, Deva Ramanan. Object detection with discriminatively trained part-based models
- Navneet Dalal, Bill Triggs. Histogram of oriented gradients for human detection

4.3 Multiview Stereo

Implement multi-view stereo that takes images from multiple viewpoints and outputs a 3D reconstruction.

- M. Goesele, B. Curless, S. Seitz. Multi-view Stereo Revisited.
- Yasutaka Furukawa, Jean Ponce. Accurate, dense, and robust multiview stereopsis

- Steven M. Seitz, Brian Curless, James Diebel, Daniel Scharstein, Richard Szeliski. A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms.

Related papers:

- Daniel Scharstein, Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms
- Sudipta N. Sinha, Daniel Scharstein, Richard Szeliski. Efficient High-Resolution Stereo Matching using Local Plane Sweeps
- Heiko Hirschmuller. Stereo processing by semiglobal matching and mutual information

4.4 Optical flow

Implement Sun et. al's Classic++ and experiment with different forms of smoothing and regularization, and explain your findings. You can evaluate your model on toy examples and also the Middlebury optical flow dataset.

- Deqing Sun, Stefan Roth, and Michael J. Black. A Quantitative Analysis of Current Practices in Optical Flow Estimation and the Principles Behind Them.

4.5 Texture synthesis

Implement PatchMatch and use it for texture synthesis

- Barnes, Connelly, et al. "PatchMatch: A randomized correspondence algorithm for structural image editing." ACM Transactions on Graphics-TOG 28.3 (2009): 24.
- Darabi, Soheil, et al. "Image melding: combining inconsistent images using patch-based synthesis." ACM Trans. Graph. 31.4 (2012): 82.

4.6 Segmentation and edge detection

Implement algorithms for edge detection and segment objects in images.

- Piotr Dollr, C Lawrence Zitnick. Fast edge detection using structured forests
- Shi, Jianbo, and Jitendra Malik. Normalized cuts and image segmentation.

Also read:

- Pablo Arbelaez, Michael Maire, Charless Fowlkes, Jitendra Malik. Contour detection and hierarchical image segmentation
- Felzenszwalb, Pedro F., and Daniel P. Huttenlocher. Efficient graph-based image segmentation.

4.7 Real-time hand pointing direction

Make a real-time application that estimates hand pointing direction and lets you draw by pointing at different parts of the screen. Training: the system moves a ball around all over the screen, and you track it by pointing at it with your finger. Test/run phase: system uses a cnn-feature-based image similarity metric to interpolate between the nearest retrieved images during the training phase to estimate a screen position where your finger is pointing.

- See 4.2, object detection papers.
- radial-basis function interpolation paper for graphics:
  `http://www.dtic.mil/dtic/tr/fulltext/u2/a259337.pdf`

4.8 Inference and optimization

Compare different algorithms for inference in computer vision, either in speed or performance.

- Szeliski et. al. A comparative study of energy minimization methods for markov random fields with smoothness-based priors.

Related papers:

- Kappes et. al. A Comparative Study of Modern Inference Techniques for Discrete Energy Minimization Problems.
- Vladimir Kolmogorov, Ramin Zabih. What energy functions can be minimized via graph cuts?

4.9 3-d print a Martian rock.

Go to a JPL web page, (for example `http://mars.jpl.nasa.gov/MPF/mpf/stereo-arc.html`), Find a stereo image of the rocks or landscape that you want to 3-d print. Use a stereo algorithm to reconstruct the 3-d shape. You'll want this to be as veridical and artifact-free as possible. You may want to use several different algorithms and average or combine the results. You may want to combine shape-from-shading with stereo. Use camera meta-data (focal length, etc) and any other visual cues to calculate how the scale of your 3-d printed rock relates to the veridical scale of the Martian rock. We'll pay or reimburse you the costs of the 3-d printing up to $200 per group (keep receipts).

- See 4.3, stereo papers.

What do you do for the part of the rock or landscape you don't see? Some options are to (a) make hidden sides have a simple geometric shape, ie, planar, spherical or elliptical, or (b) use example-based texture synthesis methods, shape-texture synthesize a plausible Martian rock shape for the missing data.

4.10 Invent new object category

Propose or invent a new object category, and make a detector for objects of that category. The detector should be as robust as possible to variations in lighting, scale, orientation, occlusion, etc.

Example new object categories could be: a person wearing a sauce-pan as a hat, or a mime, or a cube touching a sphere, etc.

- See 4.2, object detection papers.

4.11 Image Statistics

Pick one of these projects:

- Zoran and Weiss present a simple generative model for image patches, in which they model the probability of a patch with a Gaussian Mixture Model). Analyze the model that was learned (e.g. what do the covariance matrices look like? what happens when you train the model on a toy world, and use it to denoise real images?)
- Study the statistics of motion. Try applying (and extending) models that have been proposed for single images. The input could be raw video, optical flow, Fourier coefficients, etc. A recent example of this is Rosenbaum et. al., which models optical flow using a Gaussian Mixture Model.

The reading is the same for both projects:

- Daniel Zoran, Yair Weiss. Natural images, Gaussian mixtures and dead leaves
- Dan Rosenbaum, Daniel Zoran, Yair Weiss. Learning the Local Statistics of Optical Flow.