Final Projects Proposals

Computer Vision 16-720 Fall 2009

October 14, 2009

The projects listed below are suggested projects. They have been selected because they all correspond to classical approaches in the field. You can choose your own topic, but need prior approval from the Instructor.

- 30 minute presentation at the end of semester
- 1-3 papers provided for each topic below
- What is expected:
  - Overview of problem
  - Implementation of a particular idea from the papers
  - Examples (videos, Matlab examples, results on toy examples.....
  - Conclusion/limitations/problems
- What you should not do:
  - Regurgitate the papers (not useful)
  - Go through every detail of the derivations (no time and usually not useful)

Important

The papers below are provided as representative examples of the work in each area. It is very important that you check the home page of the author and of the associated lab, which contains very often a number of additional resources (videos, related papers, presentations, example code, etc.). Many of the papers are rather hard to read (or outright mysterious!) out of context and it is a good idea to use these additional resources.

In addition, tutorials, additional references can be retrieved from the usual sources:

- CVOnline
- Vision Home page
- IEEE Explore (from 128. machines only)
- Books listed in class

Important: For various reasons, many of the pdf links listed below can be accessed only from a CMU machine (i.e., with IP 128.2...) or remotely through VPN.

A SHADING, REFLECTANCE MODELS, COLOR

A.1 Shape from Shading from Examples

Reconstruction of general materials (varying albedo, general BRDFs, etc.) from a collection of training exemplars (breifly shown in class).

- Aaron Hertzmann, Steven M. Seitz. Example-Based Photometric Stereo: Shape Reconstruction with General, Varying BRDFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 8, pp. 1254-1264, August 2005. pdf
- Aaron Hertzmann, Steven M. Seitz. Shape and Materials by Example: A Photometric Stereo Approach. Proc. IEEE CVPR 2003. Madison, WI. June 2003. Vol. 1. pp. 533-540. pdf

A.2 Shading Models and Recognition

Two classical papers on shading models and their use in recognition. The key result is that the set of all images of an object under all possible illumination conditions is a low-dimensional subspace. This property is used in recognition applications.

- What is the Set of Images of an Object Under All Possible Lighting Conditions? P. Belhumeur, D. Kriegman. International Journal of Computer Vision, 28(3), 1998. pdf
- Illumination Cones for Recognition Under Variable Lighting: Faces. A. Georghiades, D. Kriegman, P. Belhumeur. IEEE Conf. on Computer Vision and Pattern Recognition, 1998. pdf

A.3 Color Constancy

The first paper is a classic that combines several classical approaches to color constancy into a single, straightforward framework. This follows closely the derivation sketched out at the end of the class notes on color. The second paper develops further one approach based on a probabilistic model.

- Color by correlation: a simple, unifying framework for color constancy. Finlayson, G.D.; Hordley, S.D.; HubeL, P.M.; Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume 23, Issue 11, Nov. 2001 Page(s):1209 - 1221 pdf
- Color constancy using KL-divergence Rosenberg, C.; Hebert, M.; Thrun, S.; Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on Volume 1, 7-14 July 2001 Page(s):239 - 246 vol.1 pdf

A.4 Haze Removal

Physics based vision algorithms for removing haze from images.

- Single Image Haze Removal Using Dark Channel Prior. K. He, J. Sun, X. Tang. In CVPR 2009. pdf
- Single image dehazing. R. Fattal. In SIGGRAPH 2008. pdf

B FILTERING, FEATURE EXTRACTION, SCALE-SPACE

B.1 Texture Classification

Classification of textures by using affine-invariant detectors (i.e., generalization of the scale-invariant detectors discussed in class). The idea is that the use of these affine-invariant features yields better robustness to changes in geometric and photometric variations.

- A sparse texture representation using local affine regions. Lazebnik, S.; Schmid, C.; Ponce, J.; Pattern Analysis and Machine Intelligence, IEEE Transactions on. Volume 27, Issue 8, Aug. 2005 Page(s):1265 - 1278 pdf
- Affine-invariant local descriptors and neighborhood statistics for texture recognition. Lazebnik, S.; Schmid, C.; Ponce, J.; Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on. 2003 Page(s):649 - 655 vol.1 pdf

B.2 Texture Classification

An approach similar to the "texton" approach used in HW2.

- Varma, M. and Zisserman, A. A statistical approach to texture classification from single images. International Journal of Computer Vision: Special Issue on Texture Analysis and Synthesis, to appear in 2005. pdf
- Varma, M. and Zisserman, A. Classifying Images of Materials: Achieving Viewpoint and Illumination Independence. Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark (2002). pdf

B.3 Scale-Invariant Representations

Another very popular way of extracting scale-invariant regions and features. This one is based on first-order derivatives (unlike the Laplacian-based technique described in class). The paper includes applications to tracking and recognition.

- Scale, Saliency and Image Description. Timor Kadir and Michael Brady. International Journal of Computer Vision. 45 (2):83-105, November 2001. pdf

B.4 Scale-Invariant Representations

Two other (related) approaches to extracting geometrically invariant regions with applications to wide-baseline stereo correspondence. These approaches are based primarily on the local distribution of intensity in the image (instead of on first or second derivatives).

- MSER: J.Matas, O. Chum, M. Urban, and T. Pajdla, Robust wide baseline stereo from maximally stable extremal regions. In BMVC p. 384-393, 2002. pdf
- IBR & EBR: T.Tuytelaars and L. Van Gool, Matching widely separated views based onaffine invariant regions. In IJCV 1(59):61-85, 2004. pdf

C CAMERA GEOMETRY AND CAMERA CALIBRATION

C.1 Geometry from a single camera

A nice application of the fundamental concepts in camera geometry. The paper shows how to recover quantitative geometric information from a single (e.g., for forensic applications). The first paper is a complete journal version. The second one is an earlier conference version of related ideas. The third paper estimates ground layer and vehicle ego-motion by using planar motion constraint. Its a very nice application of camera geoemtry transformation.

- Criminisi, A. , Reid, I. and Zisserman, A. Single View Metrology. International Journal of Computer Vision (2000) pdf
- Criminisi, A , Reid, I. and Zisserman, A. A Plane Measuring Device. Image and Vision Computing (1999) pdf
- Qifa Ke; Kanade, T. Transforming camera geometry to a virtual downward-looking camera: robust ego-motion estimation and ground-layer detection. Computer Vision and Pattern Recognition (2003) pdf

C.2 Camera Calibration

You have seen the basic single camera calibration in class. This paper uses a sequences of planar views to enforce the multiview constraints which exist between colleniations between images.

- Malis, E.; Cipolla, R. Camera self-calibration from unknown planar structures enforcing the multiview constraints between collineations. Pattern Analysis and Machine Intelligence. Volume 24, Issue 9, Sept. 2002 Page(s):1268 - 1272 pdf

D RECONSTRUCTION FROM MULTIPLE IMAGES

D.1 Factorization

A classical extension of the factorization approach to approximate perspective + multiple moving objects.

- A Paraperspective Factorization method for Shape and Motion Recovery. C.J. Poelman and T. Kanade. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(3):206-218. 1997. pdf
- A Multi-Body Factorization Method for Motion Analysis. J. Costeira and T. Kanade. Int Journal of Computer Vision, 29(3):159-180. 1998. pdf

D.2 Visual Hull

Reconstructing 3D objects from silhouettes and contours.

- The visual hull concept for silhouette-based image understanding. A. Laurentini. In PAMI 1994. pdf.
- Image Based Visual Hulls. W. Matusik, C. Buehler, R. Raskar, S. J. Gortler, and L. McMillan. In SIGGRAPH 2000. pdf

D.3 Factorization

These two papers are applications of the factorization to articulated motion recovery, in which case, the rank of the data matrix is further constrained.

- Tresadern, P.; Reid, I. Articulated structure from motion by factorization. CVPR (2005) pdf
- Yan, J.; Pollefeys, M. A factorization-based approach to articulated motion recovery. CVPR (2005) pdf

D.4 Photo Tourism

An application of structure from motion to reconstruction of sites from web images. The second paper is one element of the SfM approach used in the first paper.

- Noah Snavely, Steven M. Seitz, Richard Szeliski, "Photo tourism: Exploring photo collections in 3D," ACM Transactions on Graphics (SIGGRAPH Proceedings), 25(3), 2006, 835-846. pdf
- Schmid, C., Zisserman, A. 1997. Automatic line matching across views. In Proc. IEEE Conf. on Computer Vision and Pattern Recognition. pdf

D.5 Auto Calibration

An award-winning paper on auto-calibration, showing how 3D reconstruction is possible assuming only zero-skew of the cameras. THe second paper investigates the background in more details.

- M. Pollefeys, R. Koch and L. Van Gool. Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters, International Journal of Computer Vision, 32(1), 7-25, 1999. pdf
- M. Pollefeys and L. Van Gool, Stratified Self-Calibration with the Modulus Constraint, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 21, No.8, pp.707-724, 1999. pdf

D.6 Stereo

The first paper describes a recent approach to stereo which is based on graph cut and energy minimization algorithms. This class of algorithms is currently among the best performing algorithms for stereo. The second paper is a broader survey of relevant energy minimization techniques. Note: Some familiarity with graph algorithms and basics of MRFs is helpful for this subject. The third paper describes a cooperative stereo algorithm where uniqueness and continuity of depth values is preserved. Occlusions are also explicitly modeled in the paper.

- Fast approximate energy minimization via graph cuts. Boykov, Y.; Veksler, O.; Zabih, R.; IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Volume 23, Issue 11, Nov. 2001 Page(s):1222 - 1239 pdf
- Computing visual correspondence with occlusions using graph cuts. Kolmogorov, V.; Zabih, R. Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on Computer Vision, Volume: 2 , 2001. pdf
- A cooperative algorithm for stereo matching and occlusion detection. Zitnick, C.L.; Kanade, T. Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume 22, Issue 7, July 2000 Page(s):675 - 684 pdf
- Accurate, dense, and robust multiview stereopsis. Y Furukawa, J Ponce. CVPR 2007. pdf

D.7 3-D Reconstruction from Multiple Cameras

Another nice approach for recovering 3D from N fixed cameras. The first paper describes a simple and general algorithm for reconstruction that, unlike conventional stereo, does not require any search for correspondences. The second paper describes a more general theory inspired by this class of approaches.

- Photorealistic Scene Reconstruction by Voxel Coloring. S. M. Seitz and C. R. Dyer, International Journal of Computer Vision, 35(2), 1999, pp. 151-173. pdf
- A Theory of Shape by Space Carving. K. N. Kutulakos and S. M. Seitz. International Journal of Computer Vision, Marr Prize Special Issue, 2000, 38(3). pdf

D.8 Structure from Motion

A classical paper on 3-D reconstruction from sequences of images. Most interesting, it includes an application of the self-calibration (metric reconstruction) approaches described in class, which enable reconstruction from uncalibrated cameras. The first paper describes an entire system

- M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, R. Koch, Visual modeling with a hand-held camera, International Journal of Computer Vision 59(3), 207-232, 2004. pdf
- J. Repko, M. Pollefeys, 3D Models from Extended Uncalibrated Video Sequences: Addressing Keyframe Selection and Projective Drift, Proc. 3DIM'05. pdf
- Towards Urban 3D Reconstruction From Video. A. Akbarzadeh et al. In 3DPVT 2006. pdf

D.9 Structure from Motion

Classical papers on how to build practical system for reconstructing 3-D models from sequences of images. The second paper focuses on the generation of virtual images from image sequences.

- A. 3D Model Aquisition from Extended Image Sequence. Beardsley, P.A., Torr, P.H.S. and Zisserman. In Proc. 4th European Conference on Computer Vision, LNCS 1065, Cambridge, pages 683-695, 1996. pdf
- Automatic 3D Model Acquisition and Generation of New Images from Video Sequences. Fitzgibbon, A.W. and Zisserman, A. In Proceedings of European Signal Processing Conference (EUSIPCO '98), Rhodes, Greece, pages 1261-1269, 1998. pdf

D.10 Mapping

Constructing 3D Maps of entire cities from video data through careful use of SfM type of tools.

- A. Akbarzadeh, J.-M. Frahm, P. Mordohai, B. Clipp, C. Engels, D. Gallup, P. Merrell, M. Phelps, S. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewenius, R. Yang, G. Welch, H. Towles, D. Nister and M. Pollefeys, Towards Urban 3D Reconstruction From Video, Proc. 3DPVT'06 (Int. Symp. on 3D Data, Processing, Visualization and Transmission), 2006. pdf
- D. Gallup, J.-M. Frahm, P. Mordohai, Q. Yang and M. Pollefeys, "Real-time Plane-sweeping Stereo with Multiple Sweeping Directions", International Conference on Computer Vision and Pattern Recognition (CVPR), Minneapolis, Minnesota, USA, June 2007. pdf

D.11 Pose Estimation

Calculating camera motion between two views with only five corresponding points between the images.

- An efficient solution to the five-point relative pose problem. D. Nistr. In PAMI 2004. pdf
- A Non-Iterative Algorithm for Determining All Essential Matrices Corresponding to Five Point Pairs. J. Philip. In Photogrammetric Record, 1996. pdf

D.12 Bundle Adjustment

The first paper is a detailed explanation of Bundle Adjustment, very hughly recommended reading if you are interested in nonlinear minimization. The last two papers are practical implementations of bundle adjustment for 3-D reconstruction from sequences of images.

- Bundle Adjustment - A Modern Synthesis. Triggs B, McLauchlan P, Hartley R, and Fitzgibbon A pdf
- Efficient bundle adjustment with virtual key frames: a hierarchical approach to multi-frame structure from motion. Heung-Yeung Shum; Qifa Ke; Zhengyou Zhang. IEEE Proc. Conference on Computer Vision and Pattern Recognition, 1999. pdf
- Model-Based Bundle Adjustment with Application to Face Modeling. Ying Shan, Zicheng Liu, Zhengyou Zhang. Proc. IEEE International Conference on Computer Vision, 2001. pdf

E MOTION ANALYSIS AND SEGMENTATION

E.1 Motion Estimation and Multiview Analysis

An example of a class of approach based on "subspace analysis" which can be applied to the recovery of multiple planes in motion. Note: For those who are very confortable with linear algebra, manipulating homography matrices, etc. The second paper is an earlier version of related ideas.

- Multiview constraints on homographies. Zeinik-Manor, L.; Irani, M.; Pattern Analysis and Machine Intelligence, IEEE Transactions on. Volume 24, Issue 2, Feb. 2002 Page(s):214 - 223. pdf
- Multi-frame estimation of planar motion. Zelnik-Manor, L.; Irani, M.; Pattern Analysis and Machine Intelligence, IEEE Transactions on. Volume 22, Issue 10, Oct. 2000 Page(s):1105 - 1116. pdf

E.2 Motion Segmentation

A classical approach to motion segmentation using dominant motion. The second paper is a shorter (conference) version of the first reference. The material includes elaboration on motion models discussed in class and probabilistic models using maximum likelihood interpretation of motion estimation. The third paper is a nice application, it uses occlusions to do motion segmentation.

- Compact representations of videos through dominant and multiple motion estimation. Sawhney, H.S.; Ayer, S.; Pattern Analysis and Machine Intelligence, IEEE Transactions on. Volume 18, Issue 8, Aug. 1996 Page(s):814 - 830 pdf
- Model-based 2D&3D dominant motion estimation for mosaicing and video representation. Sawhney, H.S.; Ayer, S.; Gorkani, M.; Computer Vision, 1995. Proceedings., Fifth International Conference on. 20-23 June 1995 Page(s):583 - 590 pdf
- Motion segmentation using occlusions. Ogale, A.S.; Fermuller, C.; Aloimonos, Y. Pattern Analysis and Machine Intelligence, IEEE Transactions on. Volume 27, Issue 6, Jun 2005 Page(s):988 - 992 pdf

E.3 Image Mosaicing

Two classical papers on creating mosaics from collection of images. A direct application of the motion computation part of the class with practical algorithms and cool results. Note: The first paper is longer than usual only because it contains a lot more implementation details, not because it is more involved than the other papers.

- Image Alignment and Stitching: A Tutorial. Szeliski, R. Tech Report MSR-TR-2004-92, Microsoft Research, 2004. pdf
- Panoramic Image Mosaics. H.Y. Shum, R. Szeliski. Microsoft Research Tech Report 1997. pdf
- Robust Video Mosaicing through Topology Inference and Local to Global Alignment, Sawney, Hsu, Kumar, ECCV 1998. pdf

E.4 Motion Segmentation

The first paper is a complete development of the concept of ”layers” for segmenting scenes based on motion (e.g., ”foreground”/”background” spearation). Familiarity with Bayesian classification is recommended.

- An integrated Bayesian approach to layer extraction from image sequences. Torr, P.H.S. Szeliski, R. Anandan, P. IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 23, Number 3, 2001. pdf
- Bayesian Estimation of Layers from Multiple Images. Y. Wexler, A. Fitzgibbon and A. Zisserman. Proceedings of the 7th European Conference on Computer Vision. 2002. pdf

E.5 Motion Layers

Another approach for motion layer segmentation (results were shown in class; the paper describes the details). The second paper describes an extension to deal with outliers. The third paper discusses initialization and rank detection issues.

- Qifa Ke and Takeo Kanade, ”A Subspace Approach to Layer Extraction”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2001), Volume I, pages 255-262, Hawaii, Dec. 2001. pdf
- Qifa Ke and Takeo Kanade, ”A Robust Subspace Approach to Layer Extraction”, IEEE Workshop on Motion and Video Computing (Motion 2002), pages 37-43, Orlando, Florida, Dec. 2002. pdf
- Qifa Ke and Takeo Kanade, ”Robust Subspace Clustering by Combined Use of kNND Metric and SVD Algorithm”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2004), Washington D.C., June 2004 pdf

E.6 Motion Tracking and Motion Analysis

Two papers that describe a classical example of a system for motion analysis from image sequences. The goal of the system is to understand human activities in video. Includes motion segmentation through background substraction, tracking, and recognition.

- W4: real-time surveillance of people and their activities. Haritaoglu, I. Harwood, D. Davis, L.S. IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 22, Number 8, 2000. pdf
- Robust real-time periodic motion detection, analysis, and applications. Cutler, R.; Davis, L.S. IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 22, Number 8, 2000. pdf

E.7 Motion Segmentation and Background Substraction

More elaborate (and more recent) ways to do background substraction.

- Ahmed Elgammal, David Harwood, Larry Davis Non-parametric Model for Background Subtraction. pdf
- Anurag Mittal Nikos Paragios Motion-Based Background Subtraction using Adaptive Kernel Density Estimation CVPR 2004. pdf

E.8 Event detection/Activity recognition

Two ways of detectiong events (sitting up, waving, etc.) from videos.

- Y. Ke, R. Sukthankar, and M. Hebert. Event Detection in Crowded Videos. IEEE International Conference on Computer Vision, October, 2007. pdf
- I. Laptev, P. Prez. Retrieving actions in movies. In Proc. Int. Conf. Comp. Vis.(ICCV’07), Rio de Janeiro, Brazil, October 2007. pdf

E.9 Motion Similarity

A popular technique for estimating the similarity between the motion patterns in video blocks. This is applied to recognizing events in videos.

- Eli Shechtman and Michal Irani. Space-Time Behavior-Based Correlation OR How to Tell If Two Underlying Motion Fields Are Similar without Computing Them?
- IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 29, NO. 11, NOVEMBER 2007 pdf
- P. Matikainen, M. Hebert, R. Sukthankar, and Y. Ke. Fast Motion Consistency Through Matrix Quantization Proceedings of the British Machine Vision Conference, September, 2008, pp. 1055-1064. pdf

F TRACKING

F.1 Template/Feature Tracking

The first paper is a complete analysis of the motion recovery approaches based on the ”Lucas-Kanade” model, i.e., parameterizing the motion (u,v) by some low-dimensional model and solving by least-squares over a window. This was shown in class for constant, affine, planar, motions. The paper analyzes further the properties of this class of approaches. The second paper focuses on one detail (asked about in class): When to update the template when tracking for a long time.

- Lucas-Kanade 20 Years On: A Unifying Framework. S. Baker and I. Matthews. International Journal of Computer Vision, Vol. 56, No. 3, March, 2004, pp. 221 - 255. pdf
- The Template Update Problem. I. Matthews, T. Ishikawa, and S. Baker. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 26, No. 6, June, 2004, pp. 810 - 815. pdf

F.2 Template Tracking

Extension of the classical template tracking approach to varying illumination and non-planar shapes. Second paper extends tracking to simultaneous use of multiple trackers.

- Efficient region tracking with parametric models of geometry and illumination. Hager, G.D.; Belhumeur, P.N. Pattern Analysis and Machine Intelligence, IEEE Transactions on , Volume: 20 Issue: 10 , Oct. 1998. pdf
- Probabilistic data association methods for tracking complex visual objects. Rasmussen, C.; Hager, G.D. Pattern Analysis and Machine Intelligence, IEEE Transactions on, Volume 23, Number 6, 2001. pdf

F.3 Mean-Shift Tracking

These papers use a pupolar approach to tracking, the ”mean-shift” (also used for segmentation). The advantage of this technique is that it does not require the motion target to be restricted to a class of motions (e.g., affine) and it can deal with deformable targets. Some familiarity with kernel density estimation would help a little.

- Kernel-based object tracking. Comaniciu, D. Ramesh, V. Meer, P. IEEE Trans. Pattern Anal. Machine Intell , Vol. 25, No. 5, 2003. pdf
- Real-time tracking of non-rigid objects using mean shift. D. Comaniciu, V. Ramesh, P. Meer. Proc. IEEE Computer Vision and Pattern Recognition Conference. 2000. pdf

G SEGMENTATION

G.1 Searching Through the Space of Segmentations

In these paper, segmentation is presented as a search through the space of segmentations. In the first approach, a two stage segmentation approach is utilized where images are first oversegmented into superpixels, and then a linear classifier is trained to group together superpixels. In the DDMCMC paper, Markov Chain Monte Carlo is used to search through the space of segmentations. This paper is rather difficult to understand and can be considered a projects in its own right.

- Learning a Classification Model for Segmentation. Xiaofeng Ren and Jitendra Malik, in ICCV ’03, volume 1, pages 10-17, Nice 2003. pdf
- Image Segmentation by Data-Driven Markov Chain Monte Carlo, Z.W. Tu and S.C. Zhu, IEEE Trans on Pattern Analysis and Machine Intelligence, vol.24, no.5, pp. 657-673, May, 2002 pdf

G.2 Image Segmentation and Image Retrieval

Image segmentation using EM techniques and its application to content-based image retrieval. The second paper is an earlier (and easier to read) version. Basic understanding of expectation-maximization algorithms is useful..

- Blobworld: image segmentation using expectation-maximization and its application to image querying. Carson, C. Belongie, S. Greenspan, H. Malik, J. IEEE Trans. Pattern Anal. Machine Intell , Vol. 24, No. 8, 2002. pdf
- Color- and texture-based image segmentation using EM and its application to content-based image retrieval. Belongie, S.; Carson, C.; Greenspan, H.; Malik, J. Sixth International Conference on Computer Vision, 1998. pdf

G.3 Segmentation for Recognition

Another view of the segmentation problem, with application to extracting human shapes from images. Recovering human body configurations: combining segmentation and recognition

- Mori, G.; Xiaofeng Ren; Efros, A.A.; Malik, J.; Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. Volume 2, 27 June-2 July 2004 Page(s):II-326 - II-333 Vol.2 pdf
- G. Mori, Guiding Model Search Using Segmentation, IEEE International Conference on Computer Vision, 2005. pdf

G.4 Graph-Based Image Segmentation

Yet another segmentation approach based on graph algorithms. The second paper compares the different segmentation algorithms and introduces a way to combine the F&H approach with other approaches, such as mean shift.

- P. Felzenszwalb and D. Huttenlocher. Efficient Graph-Based Image Segmentation. International Journal of Computer Vision, Vol. 59, No. 2, September 2004. pdf
- A Comparison of Image Segmentation Algorithms. C. Pantofaru and M. Hebert. tech. report CMU-RI-TR-05-40, Robotics Institute, Carnegie Mellon University, September, 2005. pdf

G.5 Video Matting

One application of the concept of segmentation is "matting", in which one designates a small part of the image (through a mouse stroke, e.g.,) as being the background and another one as being the foreground. The segmentation algorithm uses the initial information to extract the foreground object from the background. This is a very popular topic with many applications in computer graphics and user interfaces and image editing.

- Y. Chuang and A. Agarwala and B. Curless and D. Salesin and R. Szeliski. Video matting of complex scenes. ACMGraphics. Vol.21, No. 3. 2003. pdf+examples
- J. Sun and J. Jia and C.-K. Tang and H.-Y. Shum. Poisson matting. Vol. 23, No. 3. 2004. pdf

G.6 Interactive Segmentation

A related idea: Use a little bit of manual input to help segmenting foreground object from background. Very popular also for image editing and graphics (cool examples in the papers, by the way).

- C. Rother and V. Kolmogorov and A. Blake. "grabcut": interactive foreground extraction using iterated graph cuts. ACMGraphics. Vol. 23, No. 3, 2004. pdf

- Y. Li and J. Sun and C.-K. Tang and H.-Y. Shum. Lazy snapping. ACMGraphics. Vol. 23, No. 3. 2004. pdf
- Y. Boykov and M. P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. Proc. ICCV. 2001. pdf

G.7 Spectral Matting

Another very popular technique for interactive segmentation.

- A. Levin, A. Rav-Acha, D. Lischinski. Spectral Matting. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Minneapolis, June 2007 pdf extended TR
- A. Levin, A. Rav-Acha, D. Lischinski. Spectral Matting. IEEE Trans. Pattern Analysis and Machine Intelligence, Oct 2008. pdf

G.8 Feature selection for tracking

Improvements on finding and selecting features for tracking.

- R.Collins, Y.Liu and M.Leordeanu, "On-Line Selection of Discriminative Tracking Features," IEEE Trans Pattern Analysis and Machine Intelligence (PAMI), Vol 27(10),October 2005, pp.1631-1643. pdf
- Robust Object Tracking with Regional Affine Invariant Features. S. Tran and L. Davis. In ICCV 2007. pdf

H RECOGNITION

H.1 Photometric Invariants

Matching images using local features that are invariant by rotation, translation, and scale. An important approach lately, based on extensions of the Harris detector.

- Indexing based on scale invariant interest points. The second paper is here for historical context since it is quite old. Mikolajczyk, K.; Schmid, C. Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on, Volume: 1 , 2001. pdf
- Local grayvalue invariants for image retrieval. Schmid, C.; Mohr, R. Pattern Analysis and Machine Intelligence, IEEE Transactions on , Volume: 19 Issue: 5 , May 1997. pdf

H.2 Shape Matching

Another way to match shapes using invariant shape descriptors based on local distributions of edges. Second paper discusses connections between recognition, grouping and segmentation.

- Shape matching and object recognition using shape contexts. Belongie, S. Malik, J. Puzicha, J. Pattern Analysis and Machine Intelligence, IEEE Transactions on , Volume 24, Number 4, 2002. pdf
- Visual grouping and object recognition. Malik, J. Image Analysis and Processing, 2001. Proceedings. 11th International Conference on Computer Vision, 2001. pdf

H.3 Shape Matching

Two other different, but related, papers an shape matching using quadratic programmic techniques

- A Berg, T Berg, J Malik, Shape Matching and Object Recognition using Low Distortion Correspondences, CVPR 2005. pdf
- Alternative approach: M. Leordeanu and M. Hebert, A Spectral Technique for Correspondence Problems using Pairwise Constraints, ICCV 2005. pdf
- Application to recognition and weakly-supervised learning: M. Leordeanu, M. Hebert, and R. Sukthankar. Beyond Local Appearance: Category Recognition from Pairwise Interactions of Simple Features. Proc. CVPR, June, 2007. pdf

H.4 SURF

Speeded Up Robust Features (SURF) is like SIFT, but is a faster way to calculate scale and rotational invariant features.

- SURF: Speeded Up Robust Features. H. Bay, A. Ess, T. Tuytelaars, L. V. Gool. In CVIU, 2008. pdf
- SURF: speeded up robust features. H. Bay, T. Tuytelaars, L. Van Gool, In ECCV 2006. pdf

H.5 Pictorial Structures

Another classic approach based on matching image parts and representing their relations, with applications to recognizing and tracking human shapes in images. The second paper generalizes some aspects of the initial formulation to make it applicable to braoder recognition problem. Warning: Only for people already familiar with graphical models, belief propagation and related topics (e.g., from the machine learning class).

- P. Felzenszwalb and D. Huttenlocher. Pictorial Structures for Object Recognition. International Journal of Computer Vision, Vol. 61, No. 1, January 2005. pdf
- Spatial Priors for Part-Based Recognition using Statistical Models. P. Felzenszwalb; D. Crandall; and D. Huttenlocher. IEEE Conference on Computer Vision and Pattern Recognition, 2005 pdf

H.6 Constellation models

An approach based on recognizing image parts and their relations, extracted from training data. Uses scale-invariant features and other concepts from earlier in class. The second paper is an older paper in which some of the key ideas were introduced. Warning: For those who have taken the machine learning class or equivalent.

- R. Fergus, P. Perona, and A. Zisserman. Object Class Recognition by Unsupervised Scale-Invariant Learning. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2003. pdf
- M. Weber, M. Welling and P. Perona. Unsupervised Learning of Models for Recognition. Proc. 6th European Conference Computer Vision (ECCV) Dublin, Ireland, 2000 June. pdf

H.7 Human Detection

The first paper describes one of the most popular approach at the moment for detecting humans in images, based on histograms of gradients (HoGs) descriptor. The second paper is a related implementation.

- Navneet Dalal, Bill Triggs, Cordelia Schmid. Human detection using oriented histograms of flow and appearance European Conference on Computer Vision - 2006. pdf
- Qiang Zhu, Shai Avidan, Mei-Chen Yeh, and Kwang-Ting Cheng. Fast Human Detection Using a Cascade of Histograms of Oriented Gradients. pdf

H.8 Deformable Part Model

Models objects as a probabilistic set of deformable, spatial relationships between parts of the object. One of the most successful object recognition algorithms for the PASCAL Challenge.

- A Discriminatively Trained, Multiscale, Deformable Part Model. P. Felzenszwalb, D. McAllester, D. Ramanan. In CVPR 2008. pdf
- Histograms of oriented gradients for human detection. N Dalal, B Triggs. In CVPR 2005. pdf

H.9 SIFT Flow

Based on optical flow, it uses SIFT descriptors calculated at each pixel to describe images and measure similarity between images.

- SIFT flow: dense correspondence across different scenes. C. Liu, J. Yuen, A. Torralba, J. Sivic, W. T. Freeman. In ECCV 2008. pdf
- Object recognition by scene alignment. B. Russell, A. Torralba, C. Liu, R. Fergus, W. T. Freeman. In NIPS 2007. pdf

H.10 Efficient Sliding Window

More intelligent ways of performing and optimizing sliding windows algorithms (commonly used for object detection).

- Beyond Sliding Windows: Object Localization by Efficient Subwindow Search. C.H. Lampert, M. B. Blaschko and T. Hofmann. In CVPR 2008. pdf

H.11 Recognition in Video

A nice application of all the concepts of invariant region extraction, SIFT descriptors, clustering, etc. to the problem of extracting object descriptions from video in an unsupervised manner.

- Sivic, J. and Zisserman, A. Video Data Mining Using Configurations of Viewpoint Invariant Regions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC (2004) pdf
- Sivic, J. and Zisserman, A. Video Google: A Text Retrieval Approach to Object Matching in Videos. Proceedings of the International Conference on Computer Vision (2003) pdf

H.12 Face Detection

One of the best-performing face detector based on local statistics of wavelet coefficients. Warning: Good understanding of image processing (wavelets) and machine learning (bayes classifiers, boosting) is required for this paper.

- Object Detection Using the Statistics of Parts. H. Schneiderman and T. Kanade. International Journal of Computer Vision, 2002. pdf

H.13 Using (affine) Spatial Constraints

This paper addresses recognition of objects in image with an emphasis on representing and using the 3D spatial structure of the object. This is an excellent (and relatively strightforward) exercise in integrating the concepts of invariant region detection (and Harris detector, etc.) with the concepts of affine factorization.

- Object modeling and recognition using local affine-invariant image descriptors and multi-view spatial contraints. F. Rothganger, Svetlana Lazebnik, Cordelia Schmid, Jean Ponce. International Journal of Computer Vision. To appear - 2005 pdf

H.14 Indexing in very large feature databases

Recognition reduced to the problem of designing efficient data structures for indexing in a huge training set.

- D. Nistr and H. Stewnius, Scalable Recognition with a Vocabulary Tree CVPR 2006. pdf

H.15 Combining recognition and segmentation

If we knew where the object is, we could segment it accurately; if we could produce a perfect segementation, we could recognize objects easily. These papers address this vicious circle and propose a way to combine segmentation and recognition. Note: There are 4 papers below in reverse chrnological order, but they are short (conference) papers with substantial overlap.

- Eran Borenstein, Shimon Ullman. Class-Specific, Top-Down Segmentation. ECCV 2002 pdf
- Eran Borenstein, Shimon Ullman: Learning to Segment. ECCV 2004 pdf
- E. Borenstein, E. Sharon, S. Ullman, Combining Top-Down and Bottom-Up Segmentation, Proceedings IEEE workshop on Perceptual Organization in Computer Vision, IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, June 2004. pdf
- Borenstein and Malik, Shape Guided Object Segmentation, CVPR 2006. pdf

H.16 Object Discovery

These papers are about object discovery (unsupervised learning) in images. Utilizing the idea of latent topic modeling from the statistical text processing literature, these papers model object categories as latent topics. Images are represented as bags of topics, and topics are represented as bags of words. These techniques use pLSA and/or LDA to automatically discover objects in images, and require a good amount of machine learning knowledge. The second paper uses the concept of multiple segmentations to discover segments which correspond to objects.

- Discovering Objects and their Location in Images. Josef Sivic, Bryan Russell, Alexei A. Efros, Andrew Zisserman, Bill Freeman. In ICCV 2005 pdf
- Using Multiple Segmentations to Discover Objects and their Extent in Image Collections. Bryan Russell, Alexei A. Efros, Josef Sivic, Bill Freeman, Andrew Zisserman. In CVPR 2006 pdf

H.17 Human Action Recognition

Covers descriptors in 3D (considering time as a 3rd dimension) and using them to recognize actions in videos.

- Learning realistic human actions from movies. Ivan Laptev, Marcin Marszalek, Cordelia Schmid and Benjamin Rozenfeld. In CVPR 2008. pdf
- Automatic Annotation of Human Actions in Video. O. Duchenne, I. Laptev, J. Sivic, F. Bach and J. Ponce. In ICCV 2009. pdf

H.18 Textons

More advanced ways of using filter banks/textons to perform object/scene recognition.

- TextonBoost: Joint Appearance, Shape and Context Modeling for Multi-Class Object Recognition and Segmentation. J. Shotton, J. Winn, C. Rother, A. Criminisi. In ECCV 2006. pdf
- Semantic Texton Forests for Image Categorization and Segmentation. J. Shotton, M. Johnson, R. Cipolla. In CVPR 2008. pdf

H.19 Context

Intuitively, knowing more about a scene can influence the probabilities of an object occuring in an image. This can be exploited to

- Contextual priming for object detection. A. Torralba. In IJCV 2003. pdf
- Object detection and localization using local and global features. K. Murphy, A. Torralba, D. Eaton, W. T. Freeman. In Toward Category-Level Object Recognition, 2006. pdf
- An Empirical Study of Context in Object Detection. Divvala, S.K., Hoiem, D., Hays, J.H, Efros, A.A., Hebert, M. CVPR 2008. pdf
- Putting Objects in Perspective. D. Hoiem, A.A. Efros, and M. Hebert. In CVPR 2006. pdf

H.20 Recovering Surface Layout

Classification of parts of an image as ground/wall/sky using just one image.

- Recovering Surface Layout from an Image. D. Hoiem, A.A. Efros, and M. Hebert. In IJCV 2007. pdf
- Contextual Classification with Functional Max-Margin Markov Networks. D. Munoz, J. A. Bagnell, N. Vandapel, M. Hebert. In CVPR 2009. pdf

H.21 Query Expansion

In text and image retreival, we can increase the number of relevant of results by researching using information returned from the original query.

- Total Recall: Automatic Query Expansion with a Generative Feature Model for Object Retrieval. O. Chum, J. Philbin, J. Sivic, M. Isard, A. Zisserman. In ICCV 2007. pdf
- Object retrieval with large vocabularies and fast spatial matching. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. In CVPR 2007. pdf