
CMPT888: Human Activity Recognition

Instructor: Greg Mori

Summer 2010

Syllabus DRAFT

Overview

In this course we will study approaches to tackle the problem of human activity recognition. The grand goal in this area of computer vision is to build systems that can automatically find human figures in either images or video sequences and determine what action they are performing. The recognition of human figures and actions is an important problem, with many practical applications. This technology is directly applicable to human-computer interaction, image and video retrieval and search, security and surveillance, video motion capture and automated vehicle driver assistance systems.

This seminar-style course is intended for students who wish to learn about the state of the art in vision-based human activity recognition. We will cover the major approaches in the literature, for the most part by reading and discussing research papers. Students will be expected to have, or obtain, the background knowledge in computer vision and machine learning that will be needed to understand these papers. CMPT 726 (Machine Learning) and a computer vision / image processing course would provide useful, though not essential, background knowledge. If you are concerned about your readiness for this course, please ask.

Administrivia

Lectures: Mondays 14:30-16:20, Wednesdays 14:30-15:20 in TASC1 8002

Greg's office hours: Thursdays 13:00-14:00 in TASC1 8007

Course website:

<http://www.cs.sfu.ca/~mori/courses/cmpt888/summer10>

Grading scheme

- 10% Class participation: Students will be expected to participate in discussions of the papers presented.

- 10% Reading assignments: Students will be expected to submit summaries of the one or two papers assigned each week.
- 10% Paper presentation: Each student will be expected to present one paper in class. A list of papers will be on the course website, each student should sign up to present one.
- 10% Assignment: There will be one programming assignment.
- 60% Project (5 proposal, 15 presentation, 40 report): The main component of this course is a substantial project, which may be done individually or in small groups. Students will give a presentation in the last week of classes, and submit a written report (4-6 pages).

List of topics (subject to change)

- Week 1 (May 12): Administrivia, intro
- Week 2 (May 17, 19): Preliminaries. Human detection [VJS03, DT05, FMR08], background subtraction [SG99, TKBM99], optical flow [BFB94].
- Week 3 (May 26): Motion templates [EBMM03]
- Week 4 (May 31, June 2): Greg at conference. Student paper presentations on motion templates [BD01, ZMI01, SI05, BI05, KSH07]
- Week 5 (June 7, 9): Local feature video representations [SLC04]
- Week 6 (June 14, 16): Greg at conference. Student paper presentations on local feature video representations [RYS02, DRCB05, JSWP07, MOS09]
- Week 7 (June 21, 23): Unsupervised and weakly-supervised methods [NWFF08, LMSR08, WMG09]
- Week 8 (June 28, 30): Temporal models [YOI92, BW97, LXG09, WMNG08, WM09a]
- Week 9 (July 5, 7): Human pose estimation and pose retrieval [RF03, FMZ09]
- Week 10 (July 12, 14): Discriminative methods [WM08, WM09b, FM08]
- Week 11 (July 19, 21): Human actions in still images [WJD⁺06, ICCS09, YWM10, YFF10]
- Week 12 (July 26, 28): Student paper presentations [BGS⁺05, GSSD09, CSS09, XG06, LJD09]
- Week 13 (Aug. 4, 9): Project presentations

Assignment dates

- A1: out week 2, in week 5
- Project proposal: out week 5, in week 9
- Project report: due Aug. 16

Textbooks

No required texts. The following books have been placed on hold in the library for reference:

- D. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*
- E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*
- B. Horn, *Robot Vision*
- C. Bishop, *Pattern Recognition and Machine Learning*

References

- [BD01] A. Bobick and J. Davis. The recognition of human movement using temporal templates. *IEEE Trans. PAMI*, 23(3):257–267, 2001.
- [BFB94] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. *Int. Journal of Computer Vision*, 12(1):43–77, 1994.
- [BGS⁺05] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri. Actions as space-time shapes. In *Proc. 10th Int. Conf. Computer Vision*, 2005.
- [BI05] O. Boiman and M. Irani. Detecting irregularities in images and in video. In *Proc. 10th Int. Conf. Computer Vision*, 2005.
- [BW97] Aaron F. Bobick and Andrew D. Wilson. A state-based approach to the representation and recognition of gesture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(12):1325–1337, December 1997.
- [CSS09] W. Choi, K. Shahid, and S. Savarese. “what are they doing? : Collective activity classification using spatio-temporal relationship among people”. In *9th International Workshop on Visual Surveillance*, 2009.
- [DRCB05] Piotr Dollar, Vincent Rabaud, Garrison Cottrell, and Serge Belongie. Behavior recognition via sparse spatio-temporal features. In *VS-PETS*, 2005.

- [DT05] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn.*, 2005.
- [EBMM03] A.A. Efros, A.C. Berg, G. Mori, and J. Malik. Recognizing action at a distance. In *Proc. 9th Int. Conf. Computer Vision*, volume 2, pages 726–733, 2003.
- [FM08] A. Fathi and G. Mori. Action recognition using mid-level motion features. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn.*, 2008.
- [FMR08] Pedro Felzenszwalb, David McAllester, and Deva Ramanan. A discriminatively trained, multiscale, deformable part model. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2008.
- [FMZ09] V. Ferrari, M. Marin, and A. Zisserman. Pose search: retrieving people using their pose. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn.*, 2009.
- [GSSD09] Abhinav Gupta, Praveen Srinivasan, Jianbo Shi, and Larry S. Davis. Understanding videos, constructing plots - learning a visually grounded storyline model from annotated videos. In *CVPR*, 2009.
- [ICCS09] Nazli Ikizler-Cinbis, R. Gokberk Cinbis, and Stan Sclaroff. Learning actions from the web. In *IEEE International Conference on Computer Vision*, 2009.
- [JSWP07] H. Jhuang, T. Serre, L. Wolf, and T. Poggio. A biologically inspired system for action recognition. In *ICCV*, 2007.
- [KSH07] Yan Ke, Rahul Sukthankar, and Martial Hebert. Event detection in crowded videos. In *IEEE International Conference on Computer Vision*, 2007.
- [LJD09] Zhe Lin, Zhuolin Jiang, and Larry S. Davis. Recognizing actions by shape-motion prototype trees. In *Proc. 12th Int. Conf. Computer Vision*, 2009.
- [LMSR08] Ivan Laptev, Marcin Marszalek, Cordelia Schmid, and Benjamin Rozenfeld. Learning realistic human actions from movies. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn.*, 2008.
- [LXG09] Chen Change Loy, Tao Xiang, and Shaogang Gong. Modelling activity global temporal dependencies using time delayed probabilistic graphical model. In *ICCV*, 2009.
- [MOS09] Ramin Mehran, Alexis Oyama, and Mubarak Shah. Abnormal crowd behavior detection using social force model. In *CVPR*, 2009.
- [NWFF08] J. C. Niebles, H. Wang, and L. Fei-Fei. Unsupervised learning of human action categories using spatial-temporal words. *Int. Journal of Computer Vision*, 2008.
- [RF03] D. Ramanan and D. A. Forsyth. Automatic annotation of everyday movements. In *Advances in Neural Information Processing Systems 16*, 2003.

- [RYS02] Cen Rao, Alper Yilmaz, and Mubarak Shah. View-invariant representation and recognition of actions. *Int. Journal of Computer Vision*, 50(2), 2002.
- [SG99] Chris Stauffer and W.E.L. Grimson. Adaptive background mixture models for real-time tracking. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn.*, 1999.
- [SI05] E. Shechtman and M. Irani. Space-time behavior based correlation. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn.*, 2005.
- [SLC04] Christian Schuldt, Ivan Laptev, and Barbara Caputo. Recognizing human actions: A local svm approach. In *17th International Conference on Pattern Recognition*, 2004.
- [TKBM99] Kentaro Toyama, John Krumm, Barry Brumitt, and Brian Meyers. Wallflower: Principles and practice of background maintenance. In *Proc. 7th Int. Conf. Computer Vision*, 1999.
- [VJS03] P. Viola, M. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. In *Proc. 9th Int. Conf. Computer Vision*, pages 734–741, 2003.
- [WJD⁺06] Yang Wang, Hao Jiang, Mark S. Drew, Ze-Nian Li, and Greg Mori. Unsupervised discovery of action classes. In *CVPR*, 2006.
- [WM08] Yang Wang and Greg Mori. Learning a discriminative hidden part model for human action recognition. In *Advances in Neural Information Processing Systems (NIPS) 21*, 2008.
- [WM09a] Yang Wang and Greg Mori. Human action recognition by semi-latent topic models. *IEEE Transactions on Pattern Analysis and Machine Intelligence Special Issue on Probabilistic Graphical Models in Computer Vision*, 31(10):1762–1774, 2009.
- [WM09b] Yang Wang and Greg Mori. Max-margin hidden conditional random fields for human action recognition. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn.*, 2009.
- [WMG09] X. Wang, X. Ma, and E. Grimson. Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models. *IEEE Trans. PAMI*, 31(3):539–555, 2009.
- [WMNG08] Xiaogang Wang, Keng Teck Ma, Gee Wah Ng, and W. Eric L. Grimson. Trajectory analysis and semantic region modeling using a nonparametric bayesian model. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn.*, 2008.
- [XG06] T. Xiang and S. Gong. Beyond tracking: Modelling activity and understanding behaviour. *Int. Journal of Computer Vision*, 67(1):21–51, 2006.

- [YFF10] Bangpeng Yao and Li Fei-Fei. Modeling mutual context of object and human pose in human-object interaction activities. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn.*, 2010.
- [YOI92] J. Yamato, J. Ohya, and K. Ishii. Recognizing human action in time-sequential images using hidden markov model. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn.*, 1992.
- [YWM10] Weilong Yang, Yang Wang, and Greg Mori. Recognizing human actions from still images with latent poses. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn.*, 2010.
- [ZMI01] Lihi Zelnik-Manor and Michal Irani. Event-based video analysis. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn.*, 2001.