Object Tracking and Recognition using Background Estimation

Masayuki Yokoyama and Tomaso Poggio

The Problem: We consider the problem of motion analysis, particularly tracking and recognition of moving objects. We suppose to use inputs from a single and static camera. (But vibrations must be taken into account.) And we also suppose that the database for recognition is created from a video sequence captured on the same camera.

It has been a difficult problem to detect moving objects precisely, because of following factors.

- illuminance changes
- small motions of background (e.g. leaves, edges, oceans, etc.)
- shadows
- occlusion

Moreover, we need to solve other problems such as affine changes, especially scaling and rotation, for the accurate tracking and recognition.

Motivation: For designing an optimal circuitry for computer vision, tasks of low level vision cannot be ignored. Proposing and designing an applicable motion analysis system using both low and high level tasks helps us developing LSIs for computer vision systems.

Previous Work: In case of using a static camera, background estimation is a popular method for detecting foreground objects. There are many researches of background estimation[4]. Elgammal et. al. [1] adapts mixture of Gaussians and non-parametric model to an image, but it supposes only one Gaussian to a pixel.

Approach: Our approach for detecting objects will take following steps.

Firstly, we will extract foreground pixels from a frame. We suppose that color and intensity of a background pixel obeys one or some normal distribution. A correlation between color and intensity of a pixel can be measured from 2D histogram, whose dimensions represents hue and intensity respectively. They are obtained from HSV color space which is close to perception of a human retina. We accumulate hue and intensity values of frames to the histogram, only in case that the pixel has no velocity.We use optical flow techniques[3] to calculate velocity field, and define the smallest size of foreground objects for removing noise.

Secondly, we will cluster foreground pixels using their color, position, and velocity information. How to cluster pixels is currently under investigation. Once clusters are obtained, we can use them as inputs of tracking and recognition process.

We are currently examining mean-shift based approaches[2] for tracking objects. How to transform objects to feature space is our future work for recognizing once detected objects.

Difficulty: An accurate motion analysis without using expensive resources (e.g. large occupation of processor tasks and memory capacity) is required. So how to decrease calculation time is one of the most difficult problem, especially in case of real-time analysis. Such a problem of cost and performance often become a trade-off between them.

Impact: We suppose that the platform hardwares for our research do not have high-speed processors and much memory capacity, so it is our hope that our research promotes development of a practical motion analysis system. Mobile platform is our ideal target, whose clock frequency and memory size are rigidly restricted for its low power consumption.

Future Work: The goal of our research is not only tracking objects, but also learning detected objects and recognizing reappeared objects. We are now examining which feature space is suitable for the recognition.

Research Support: Research support for CBCL is provided by Government grants from DARPA, and the NSF. Corporate support is provided by AT&T, Central Research Institute of Electric Power Industry, Center for e-Business (MIT), Daimler Chrysler AG, Compaq, Eastman Kodak Company, Honda R&D

Co.,Ltd., ITRI, Komatsu Ltd., Merill-Lynch, Mitsubishi Corporation, NEC Fund, Nippon Telegraph & Telephone, Oxygen, Siemens Corporate Research, Inc., Sumitomo Metal Industries, Toyota Motor Corporation, WatchVision Co., Ltd., and the Whitaker Foundation.

References:

[1] D. Harwood A. Elgammal and L. Davis. Non-parametric model for background subtraction. In *ECCV*, volume II, pages 751–767, May 2000.

[2] G.R. Bradski *et al*. Computer vision face tracking for use in a perceptual user interface. *Intel Technology Journal*, Q2:1–15, 1998.

[3] D.J. Fleet J.L. Barron and S.S. Beauchemin. Performance of optical flow techniques. *IJCV*, 12(1):43–77, February 1994.

[4] A. Mittal and N. Paragios. Motion-based background subtraction using adaptive kernel density estimation. In *CVPR*, volume 2, pages 302–309, 2004.