Video Content Analysis

Session Organizers: Yoshikazu Nakajima, University of Tokyo, and Andrew Gallagher, Cornell University

In the past several decades, photography has undergone a digital revolution. In the past, capturing images was a task done to preserve a moment, or by capturing a set of images, to tell a story.

Now, billions of digital cameras are in the world, capturing billions of images per day. Many of these cameras are video-capable, capturing dozens of images per second. Combined with powerful microprocessors and algorithms to analyze the content or manipulate the video stream, many powerful applications have and will emerge. The scope of video content analysis is expanding, so this session includes portions of related research fields such as computer vision and computer-human interface.

Video analysis is being used to create 3D models of environments, to perform security surveillance or monitoring, to track humans and other objects, to augment reality, to help recommend videos based on personal preferences, and to provide new way for humans to interact with computers. For example, a jumping person can help a character avoid a pit in a video game by real-time analysis of video.

The talks in this session will focus on broad areas of video content analysis. The time-dimension of video captures the motion of objects or people over time, and how objects look from other points of view. This huge amount of information is distilled to model the structure of objects and their behaviors and motions in a scene. The speakers in this session will discuss the concepts behind the analysis, along with the opportunities for video analysis in the field.

The session will begin with a presentation from Rogerio Feris (IBM) about an attribute-based system for finding people in surveillance video. With this system, an operator could find people entering a specific building wearing a red sweater and sunglasses, for example. Next, Alex Berg (Stony Brook University) will describe video analysis for tracking specific people and recognizing their actions. Tools that are both low-level, such as optical flow for following pixels in time, and high level, such as modeling the bump-avoidance behavior of humans in a crowd, make this tracking possible. Next, Yoshinobu Hotta (Fujitsu Laboratories) will describe video retrieval for organizing and providing access to vast quantities of video with auto-tagging, low-level feature analysis such color and shape, and high-level features such as scene and camera motion analysis. Finally, Kiyoshi Kiyokawa (Osaka University) will conclude the session. His talk will describe video-based augmented reality (adding computer information to the real world) and mixed reality (mixing real and synthetic environments) applications, and will highlight the importance of understanding when the user needs certain information, and how best to present it.