Disney Research Method Automatically Edits Footage From Wearable Cameras Into Coherent Videos

Shared Focus of Attention From Multiple Cameras Identifies Most Interesting Content

Video cameras that people wear to record daily activities are creating a novel form of creative and informative media. But this footage also poses a challenge: how to expeditiously edit hours of raw video into something watchable. One solution, according to Disney researchers, is to automate the editing process by leveraging the first-person viewpoints of multiple cameras to find the areas of greatest interest in the scene.

The method they developed can automatically combine footage of a single event shot by several such "social cameras" into a coherent, condensed video. The algorithm selects footage based both on its understanding of the most interesting content in the scene and on established rules of cinematography.

"The resulting videos might not have the same narrative or technical complexity that a human editor could achieve, but they capture the essential action and, in our experiments, were often similar in spirit to those produced by professionals," said Ariel Shamir, an associate professor of computer science at the Interdisciplinary Center, Herzliya, Israel, and a member of the Disney Research Pittsburgh team.

The Disney Research Pittsburgh scientists will present their findings at ACM SIGGRAPH 2014, the International Conference on Computer Graphics and Interactive Techniques, Aug. 10-14, in Vancouver, Canada.

Whether attached to clothing, embedded in eyeglasses or held in hand, social cameras capture a view of daily life that is highly personal but also frequently rough and shaky. As more

- more –

people begin using these cameras, however, videos from multiple points of view will be available of parties, sporting events, recreational activities, performances and other encounters.

"Though each individual has a different view of the event, everyone is typically looking at, and therefore recording, the same activity – the most interesting activity," said Yaser Sheikh, an associate research professor of robotics at Carnegie Mellon University. "By determining the orientation of each camera, we can calculate the gaze concurrence, or 3D joint attention, of the group. Our automated editing method uses this as a signal indicating what action is most significant at any given time."

In a basketball game, for instance, players spend much of their time with their eyes on the ball. So if each player is wearing a head-mounted social camera, editing based on the gaze concurrence of the players will tend to follow the ball as well, including long passes and shots to the basket.

The algorithm chooses which camera view to use based on which has the best quality view of the action, but also on standard cinematographic guidelines. These include the 180-degree rule – shooting the subject from the same side, so as not to confuse the viewer by the abrupt reversals of action that occur when switching views between opposite sides.

Avoiding jump cuts between cameras with similar views of the action and avoiding very short-duration shots are among the other rules the algorithm obeys to produce an aesthetically pleasing video.

The computation necessary to achieve these results can take several hours. By contrast, professional editors using the same raw camera feeds took an average of more than 20 hours to create a few minutes of video.

The algorithm also can be used to assist professional editors tasked with editing large amounts of footage.

Other methods available for automatically or semi-automatically combining footage from multiple cameras appear limited to choosing the most stable or best lit views and periodically switching between them, the researchers observed. Such methods can fail to follow the action and, because they do not know the spatial relationship of the cameras, cannot take into consideration cinematographic guidelines such as the 180-degree rule and jump cuts.

In addition to Sheikh and Shamir, the research team included Jessica Hodgins, VP of

- more -

Disney Research and a professor of computer science and robotics at Carnegie Mellon; Hyun Soo Park, then a Ph.D. student in mechanical engineering at Carnegie Mellon University and Ido Arev, a computer science student at Herzliya and a Disney Research Pittsburgh intern.

For more information and a video, visit the project home page at http://www.disneyresearch.com/project/automatic-social-editing/.

###