The OZ Project: Observing Attention

Roberta Ferrario\textsuperscript{1}  Chiara Bassetti\textsuperscript{1}
Nicola Conci\textsuperscript{3}  Davide Conigliaro\textsuperscript{1,2}
Marco Cristani\textsuperscript{2}  Paolo Rota\textsuperscript{3}
Nicu Sebe\textsuperscript{3}  Francesco Setti\textsuperscript{1}

\textsuperscript{1}Istituto di Scienze e Tecnologie della Cognizione – CNR, Trento
\textsuperscript{2}Università degli Studi di Verona
\textsuperscript{3}Università degli Studi di Trento

March 21, 2014
Understanding whether an event has attracted the audience attention and singling out the moments that have been mostly enjoyable is a primary goal for those who manage sport and show business, in order to ameliorate presentation strategies, logistic aspects, and the like of an event.

But studying where the attention of a crowd, whose movements are constrained by the surrounding environment, is directed can be of great help in many other contexts, like public crowd management, space design, virtual and intelligent environments...

So far, such information has been inferred on the basis of the number of sold tickets and *ex-post* polls. Our proposal is to try and gather some of such information in an automatic manner, by using computer vision techniques.
Pre-evaluation Phase

- Analysis of relevant (social psychology) literature to study which are the movements that are typically performed in situations of interest.

- Elaborating on this, construction and successive refinements of a computer vision algorithm able to identify and visually highlight the relevant movements.
An Interdisciplinary, Mixed-methods Project (II)

Execution Phase

- Recording of video-sequences
- Detection \textit{at run time} of level of attention, excitement etc.
- Performance of a survey within the audience

Post-evaluation Phase

- Comparison of the results of the computer vision algorithm with those of the \textit{a posteriori} survey
- Labeling of elementary actions
- Construction of an ontological model and automatic reasoning and inference (\textit{logic})
- Comparison with video analysis (\textit{sociology}) results

- orientation (facing)
- vocalization (producing sounds other than words with mouth)
- verbalization (uttering words)
- vertical locomotion (movement of the body over the same point on the ground)
- horizontal locomotion (movement of the body from one point on the ground to another)
- gesticulation (meaningful bodily configuration based on fingers, hands, and arms movements mainly)
- manipulation (using hands to applaud or to strike, carry, throw, pull, etc.)
Where We Are So Far... the Video Sequences to be Analyzed

- All analyzed videos have been taken during the 2013 IIHF Ice Hockey U18 World Championship, Asiago (VI), April 7-13, 2013
- Two entire matches were recorded (Italy VS Norway, Italy VS Slovenia), each by two cameras, mounted frontally at a distance of about 25 meters from the spectators’ stand. Each camera was pointing at an half of the whole stand, the zoom being fixed. Therefore, for each match we have two sequences, further divided in 3 as the times of the hockey play. This resulted in 12 videos, for a total duration of about 6 hours
- All videos have been manually labeled by highlighting the main actions of the game, especially the fouls, shots and goals
- Italy VS Norway ended 1-12, while Italy VS Slovenia 3-2
Flow direction is measured in order to capture orientation, vertical and horizontal locomotion.

Entropy is measured in order to capture gesticulation and manipulation.

Such measures, if given in connection with the relevant events that have been manually tagged in the video sequences, allow to:

- distinguish fan groups belonging to different teams by automatically detecting their liveliness in different moments of the match, *even when they are merged* in the stands.
- single out the most salient events of the match, like goals, fouls or shots on goal, only by automatically detecting crowds motions on the stands.
Spectators segmentation and excitement calculation

a) an example frame of the sequence; b) spectators segmentation results, where the regions are colored considering their mean excitement level; c) dendrogram of the temporal clustering; d) zoomed images, highlighting the behavior of people of the different regions e) a frame of the match played in the considered interval.
a) two different frames of the sequence, the first extracted during the Norwegian goal, the second during the Italian goal. b) in the middle the spectators segmentation result, where the regions are colored considering their mean excitement level. Above and below zoomed images, highlighting the behavior of people in the different regions related with the goals of the different teams. c) the two goals.