Beyond The Kiss-Cam: Measuring The Fan Using Computer Vision-based Analytics

George Williams   Christoph Bregler   Ian Spiro   Christoph Bregler
Dept. of Computer Science, Courant Institute, New York University
george,chris,ian@movement.nyu.edu

Abstract

What are fans really doing at the game? When are they watching the action on the field/court, and when are they buried in their phone or tablet ? Which ads do they notice on the JumboTron, and which ones do they ignore ? Whos joining in the wave and whos not? Are they having fun ?

If you are a club trying to retain seat-buying fans, fill empty chairs, or effectively use an advertising budget, then these are important questions to answer.

Video surveillance can be a useful tool here, but in and of itself cannot scale to large stadium crowds. In this paper, we describe a method that applies vision-based techniques to surveillance footage. Our method attempts to measure stadium fan engagement and sentiment, via body gesture- and action-recognition, and to measure visual attention, via estimated gaze. We combine both machine learning algorithms and inexpensive, crowd-sourcing techniques to achieve speed and scale. Moreover, we show how its possible to preserve the anonymity of the fans during the analysis.

We experimented using crowd footage at a college basketball game and

1. Introduction

2 Related Work

3 Approach

3.1 Data Collection: Game Day

We set up 4 HD cameras at a college basketball game and positioned them above the audience at balcony level. We aimed 3 cameras at the most populated areas of the stands. The camera views did not overlap. We manually adjusted the zoom so that we captured as many fans as possible in each view, but not so far zoomed-out that each face spanned no less than around 12 pixels. Any less than 12 pixels and the performance of the software-based face detector decreases

Figure 1. Examples of Motion Signatures. The green colored signature shows how the orientation of the sparse flow features changes over time (top row), and the red-blue color coded signature shows the delta features (middle row).

significantly [REF]. We aimed the final camera at the basketball game and adjusted the zoom to capture the action at both ends of the court, including the scoreboard which hovered above the center of the court. Figure XX shows 3 of the camera views.

3.2 MOS: Motion Orientation Signatures

3.3 Face Tracking and Grid based Coarse Locality

3.4 GMM-Super-Features

3.5 Discriminant Classification

3.6 Comparison with Bag-Of-Feature Architectures

4 Multi-Modal Experiments

Figure **??** shows some example video frames from our database. In total we have 72 minutes of video data with 320×240 pixel resolution. For each subject we have at least 4 different video clips recorded at different times. Most subjects are public figures like international politicians or talk show hosts. We are currently also collecting a database of "non-famous" people.

5 Discussion

6 Acknowledgements