

Soccer event detection via collaborative multimodal feature analysis and candidate ranking

ABSTRACT

This paper presents a framework for soccer event detection through collaborative analysis of the textual, visual and aural modalities. The basic notion is to decompose a match video into smaller segments until ultimately the desired eventful segment is identified. Simple features are considered namely the minute-by-minute reports from sports websites (i.e. text), the semantic shot classes of far and closeup-views (i.e. visual), and the low-level features of pitch and log-energy (i.e. audio). The framework demonstrates that despite considering simple features, and by averting the use of labeled training examples, event detection can be achieved at very high accuracy. Experiments conducted on ~30-hours of soccer video show very promising results for the detection of goals, penalties, yellow cards and red cards.

Keyword: Soccer event detection; Sports video analysis; Semantic gap; Webcasting text