Using inertial sensors to index into video

David D Rowlands\textsuperscript{a,}\textsuperscript{*}, Mitchell McCarthy\textsuperscript{a}, Daniel A James\textsuperscript{a,b}

\textsuperscript{a}Wireless Monitoring and Applications, Griffith University, Nathan, Brisbane, 4111, Australia
\textsuperscript{b}Center of Excellence for Applied Sports Science Research, Queensland Academy of Sport, Brisbane, Australia

Accepted 02 March 2012

Abstract

Video analysis is a very important tool that is used by players, coaches and sports scientists since it is more intuitive and familiar than other analysis methods. One of the issues with using video is searching the video data to look for specific events to examine. The common technique for searching the video data involves watching the video and hand scoring it for later use or by visually searching until the event of interest appears. There have been attempts to automatically score the data using image processing techniques but these require elaborate multicamera systems coupled with complex image processing software.

Inertial sensors are cheap and readily available and have been used to great effect to analyse athlete’s performance and to detect selected events in the performance. The advantage of inertial sensors is that they can be mounted on the person or sporting equipment and can continually monitor the performance without suffering the problems of lighting, angle, and occlusion which occur in video systems.

This paper describes a technique to extract events using inertial sensors and using the timing of those events to index into synchronised video. The event type and timing information derived from the inertial sensor data can be stored and used as search keys for specific events in the video. This technique is demonstrated through a tennis visualisation system that uses strokes derived from a racquet mounted sensor as the index into the video.

© 2012 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.

Keywords: Inertial sensors; video index; tennis; stroke detection

1. Introduction

Video analysis is used in a number of different scenarios in the sporting arena. It can be used as a technique analysis tool for coaches and sports scientists or it can be used as part of the broadcast media.

* Corresponding author. Tel.: +61-7-3735-5383; fax: +61-7-3735-5384.
E-mail address: d.rowlands@griffith.edu.au.
which is more in the public domain. [1-3]. For coaches and sports scientists the video will be a more intuitive and comfortable method of analyzing the athlete since it closely analogues the activity as performed in the field. The video is also a good tool for analyzing past performances and looking for improvement in technique. For the sports broadcaster, the video is the public face of the sport. The broadcaster may wish to give highlights or replays of certain actions such as a goal in soccer or a strike in baseball [1]. However what both of these scenarios have in common is the fact that they generate large amounts of video which are interspersed by the key events of interest. This large amount of video also has repercussions on the storage requirements so it would be more efficient and useful if the key events could be extracted and stored with the extraneous video removed. Therefore the requirement is to extract these events from the video.

The current technique to extract the events is for the coach, broadcaster or documentalist is to manually sift through the video and annotate the video with an indicator of the events which typically contains a timestamp.[2][4] In broadcasting, a controller is watching the video feed as the play occurs, and marks the plays of interest for replay. Manually annotating is a time consuming and expensive task that highlights the need for automatic methods for video indexing [2-4]. Video indexing centres on the need to detect and classify certain events in the video. In a number of sports the events of interest are interspersed by significant periods of time that are not directly relevant to the event of interest. For example, if the coach is interested in the golf swing then the video will contain a number of short duration swing events coupled with other material that is not directly relevant to the swing. Tennis is another example of this where the coach may only be interested in backhands and will need to sift through a lot of other strokes to examine the ones of interest.

There have been many attempts to automatically index video based upon event detection in a number of different sports such as soccer, swimming, tennis etc [1][3][4]. Many of these systems use image analysis techniques directly off the video itself or use a multimodal approach incorporating image analysis and data from other sources such as inertial sensors [4] or other external knowledge based sources such as match reports, statistics on the match from the web or newspapers etc [1]. Drawbacks of these video techniques are that they require elaborate multi-camera systems coupled with complex image processing software to overcome any occlusions as well as any lighting issues.

Inertial sensors are cheap and readily available and have been used to great effect to analyse athlete’s performance and to detect selected events in the performance. [5-8] The advantage of inertial sensors is that they can be mounted on the person or sporting equipment and can continually monitor the performance without suffering the problems of video systems. Inertial sensors do suffer from being more invasive then just filming video since the athlete is required to wear a sensor or have a sensor mounted on the sporting equipment. However, with the continually reducing size of the inertial sensor platforms then this is becoming less of a problem.

Event extraction from the inertial sensor data can be more simple than from video analysis techniques which makes them attractive to use. For example detecting a stroke in tennis is just a threshold in the accelerometer data [3][5] whereas using the video image alone requires more complex analysis[2][3]. In order to assist the video indexing have greater relevance then it can be useful to classify the events into different event types so that the context of the event can be added for later retrieval. This can be done through video analysis, inertial sensor analysis, or multi-sensor techniques. Inertial sensors present themselves as a good candidate for indexing of events due to their relative ease of use, their availability and low cost, and require less of a complex setup than video capture systems.

This paper discusses a methodology to allow inertial sensors to be used for video indexing and the gives an example application of stroke detection in tennis.
2. Using Inertial Sensors for Video Indexing

Figure 1 shows the conceptual flow diagram for an inertial video indexing system. The flow shows the conceptual steps from the first capture of the inertial data and video, to the extraction and classification of the events and the storage of this video indexing metadata in a database or file system. Once the metadata has been stored along with the video or video fragments, it can then searched for specific event retrieval analysis and/or display in the visualization software.

2.1. Data Capture

The data capture requires a capture of the inertial data (accelerometer, gyroscopes), a time stamp for the inertial sensor measurements, and video data. It is important that both the inertial sensor measurements and the video data are synchronized. This is important so that when an event is detected then the same sample time point can refer to the relevant frame in the video.

2.2. Event Extraction and Classification

The event extraction and classification are the major component of this methodology as it is this step that defines the event boundaries in time and identifies the type of event. The extraction and the classification have been broken into two parts since they are two separate concepts. For example, in tennis the timing of the stroke can be detected but once the stroke is found it can be further classified into forehand, backhand to give greater context. The inertial data (single sensor or multi-sensor) can be scanned to determine an event and will pass that events signature to the classifier along with the timing of the event. The classifier will then use this information to further categorize the activity. For example a
peak detect can be used to detect a stroke in tennis and the gyroscope can then be used to further classify it into forehand and backhand. These steps can be as complex as required and can range from simple template matching, to statistical based modeling [4] and even extend to neural networks.

2.3. Storage

This is where the video and data is stored. It can range from a flat file system to a database system with both requiring the storage of the events and the event types along with the video. Significant reduction in the video storage can be achieved by just saving a window of video frames around the events. This can also speed up the retrieval and transmission of the video data to a remote site for analysis. It can also make the use of the database more efficient than just storing the full video file and a list of indices into the file.

2.4. Visualisation Client

The Visualisation Client will display the results of the indexed video to the coach/ broadcaster etc. It can be remote or local to the data storage with the most noticeable difference between them being the speed of video transfer to the visualisation client. It is expected that this client will be an interface to allow the searching of events and the display of events. As an example application given in section 3, the client can search for the 42nd stroke in a tennis match and display the video for just that stroke.

3. Example Application: Tennis

An example system using video indexing of tennis based upon stroke detection is shown in figure 2. Figure 2(a) shows a serve and figure 2(b) shows a forehand ground shot. The strokes are determined using inertial sensors (gyroscope and accelerometer) attached to the racquet of the player. The system will play the video and display the inertial data for a 1 second time window around the selected stroke. The system is composed of a number of component areas. A file loading area, a view control area, an inertial sensor viewing area, a stroke selection area, a video viewing area, and a stroke data area.

The file loading area allows the relevant files to be loaded. In this example the stroke detection was already performed with the timing of the stroke saved into a file. The view control area allows adjustment of the viewing parameters of the relevant inertial sensor data. The pre and post stroke viewing timings can be adjusted here. In the figure, a viewing window of 1 second was chosen with a window comprising of 2/3 second pre shot and 1/3 of a second post shot. This corresponds to 30 frames of video. The stroke selection area allows the relevant shot to be selected. Upon selection the relevant window of inertial data is loaded and the corresponding window of video data is loaded. The selected stroke can be seen as a red stalk amongst all the detected strokes in the video viewing area. The inertial sensor viewing area displays the window of inertial sensor data around the stroke. As the video is played a black vertical line sweeps across the graph indicating the current frame being displayed in the video viewing area. The video viewing area is used to display the video. The stroke data area is used to give information about the video frames under investigation and to control the playback of the video.
Fig. 2. Example tennis video indexing system showing the video and the inertial data for a 1 second time window around a selected stroke. (a) shows a serve and (b) shows a forehand. The black line indicates the current position of the video in the inertial data. The red stalk indicates the current stroke position in all the detected strokes.

In the current system the strokes are detected with a one second time window around the stroke. If just the video was saved as a 1 second window around all the strokes then this will be a large reduction in the storage required. In the example shown in figure 2 the saving will be 95%. The video has 50,320 frames, there are 93 strokes with 30 frames for each stroke (2,760 frames). Therefore the resulting ratio of stroke frames to total frames is 5% which represents a large saving in frames to be stored. It should be noted that this video was recording everything from the first stroke to the last stroke. The ratio will change if the camera was paused in intervening dead periods.

This work conformed to Griffith University’s ethical standards. (GU ethics code: ENG/11/11/HREC)

4. Conclusion

This paper has given the conceptual technique for using inertial sensor data to index into video. The requirement is that the video data and the inertial sensor data be synchronized. The technique hinges on being able to detect events in the inertial data, which are relevant to the sport under investigation. This
allows the technique to be generalized and therefore applied to any sport given a suitable event detection algorithm.

The concept was demonstrated using inertial sensors to detect stroke events in tennis and a system was created to visualize the tennis game in terms of stroke events. This had the advantage of being able to skip large amount of irrelevant video data which sped up the process of the analysis. It was also shown that if just the video around the events was stored then significant savings in video was achieved.

Overall, this paper has demonstrated a general technique and applied it to tennis to show that inertial sensors can be used to index into video with significant savings in video size and the time spent in video searching.

References