Muithu: Smaller Footprint, Potentially Larger Imprint

Dag Johansen Magnus Stenhaug Roger B. Hansen
University of Tromsø
Tromsø, Norway

Agnar Christensen Per-Mathias Høgmo
Tromsø Idrettslag
Tromsø, Norway

Technical Report 2012-72
Dept. of Computer Science
University of Tromsø

***Abstract*—We describe our experience with the Muithu sports notational analysis system, a novel digital information system in the popular sports domain. The system integrates real-time coach notations with related video sequences, and is configured with small, off-the shelf and cheap components. Muithu requires little or no human post-processing, which is in strong contrast to state-of-the art resource-intensive competing systems. Muithu also provides a novel social network for athletes and their coaches for information management and interactive e-learning experiences based on video footage. This next generation digital information system is already in operational use by a Norwegian elite soccer club, both for training and game events.**

Keywords—sport analytics; mobile notational analysis system; social network; multimedia sensor network; disruptiv technology.

I. INTRODUCTION

Digital information systems are currently being widely adopted in and around sport activities and events. One aspect is related to sport entertainment, where previous and on-going statistics and results enrich viewer experiences. To illustrate technology infusion in the sports domain, the upcoming Olympics in London this year has a Technology Operations Center staffed with 450 technicians. The organizers connect over 100 sporting venues with 16,500 IP handsets and 6,000 web conferencing clients to support in the order of billions of PCs, smart phones and tablets predicted to be connected to the event while running [1].

Another aspect is that digital information systems might provide the comparative benefit that distinguishes top athlete performance from their competitors. For instance, video footage of sport athlete performance is widely used to provide accurate and objective information to top athletes and their coaches. Digital notational analysis systems can be used complementary by providing a list of measurable parameters based on the observed performance. Such quantitative parameters are relative to the sport and activity at focus, and example events include that a specific soccer player passes the ball into the opponent end-zone, or that a rugby player is being tackled. Some notational analysis systems require manual annotation of such events [2] while others are more semi-automated [3].

We are in particular interested in sports with non-linear flowing performance, which makes it harder for automated analysis. Soccer is an example of a flowing game, and many premier league soccer clubs already use this type of systems. This includes Barcelona FC, Inter Milan, New York Red Bulls, Manchester United, Liverpool FC, Chelsea FC, and Arsenal FC.

Unfortunately, there are limitations and problems with several of these notational analysis systems. For instance, humans with domain expertise still need to manually identify performance events of a certain complexity. Head coaches can best determine perceived performance against complex, yet often detailed and dynamic strategic goals, but rarely have capacity for this tedious and time-consuming analytics process. As a compromise, a collection of static cameras are wired together with computers running analytics software, but where a set of additional human experts also needs to be in the critical analytics path. With such a large footprint and operational cost, applicability of existing systems is primarily for main events like games, not for day-to-day training sessions and practices.

We have developed Muithu, a complementary, light-weight video and cellular phone based notational analysis system for this purpose. Probably contrary to what to expect for a small footprint system, have we added the head coaches to the notational analysis loop. Our main conjectures for such a digital information system include that it has to be non-invasive for the head coaches, yet independent of a large group of human analysts in the back-end. The system also has to be close to fully automated, mobile, scalable, with a minimalistic footprint and low operational cost, and, last but not least, applicable in close to real-time. This set of properties is, if we may say so, novel and might have a large imprint if possible to retrofit them all into a single, applicable system.

The rest of this paper is structured as follows. In section II, we detail the properties of Muithu. The guiding principles for engineering the system are explained in Section III. Section IV presents the Muithu architecture and design. We divide our experience into two categories and present each in Section V. Section VI relates to other work, while Section VII concludes.

II. MUITHU SYSTEM PROPERTIES

Athletes spend much more time preparing for competition than actually competing. Hence, a key property of Muithu is to support the improvement of athlete performances through video analysis and visual feedback from training practices. This might come as a surprise to illiterate sport analysts, but the complexity and labor-intensity of existing digital information systems in this domain make them hard to put to everyday use. Simultaneously, we have also attempted to build our digital

Muithu translates to *memory* in the indigenous Arctic North-Saami Language. Muithu also plays on some oral resemblance with "MyTube".

information management system so that it can be useful for game analytics purposes.

We have distilled and specified a set of properties working alongside elite soccer coaches and professional athletes. The main goal has been to develop a practical system for athlete coaching, so we have iterated over a series of prototypes developed and respectively used them in practice on the training fields. This has resulted in the following mature list of requirements for the Muithu system.

Notations and related videos: The system shall capture coach-perceived performance events during practice, but with minimalistic human effort. A short video sequence correlating to each annotated event must be captured, stored, correlated with meta-data, and prepared for viewing in a coaching context. The net effect is a system that continuously captures videos of the sports field during practice, but automatically filters out video sequences of all key episodes and correlates them with their respective analysis notations.

High recall and precision: Recall is an indicator for how many of the important events triggering a coach response during a session are recollected. This is known to be a serious problem for sport coaches with a recall in soccer evaluated to be in the order of 0.3 [4]. Hence, coaches can only recollect 30% of incidents that determine successful performance, but our system must guarantee that all events explicitly annotated can be revoked in retrospect.

Combining high recall with high precision is also important to avoid information overload problems. Hence, the goal is to achieve close to recall = 1 (100%) and precision = 1 (100%) simultaneously.

Automated: The system shall be fully operational without an additional analyst group needed to post-browse large collections manually. Since we add the coaches to the notational loop in real-time, the system should to a large extent avoid interference with their mode of operandi. The only input required by them is thus a set of measurable performance markers defined prior to the practice. Next, during practice should event notations be explicitly captured with minimalistic and non-intrusive efforts; together with coaches have we defined the quality of service parameter of an average event notation process to be acceptable if it is in the order of 5 seconds or less.

Mobile: The system must be mobile and easily portable to different spatial settings, not only pre-calibrated at a home arena as state-of-the art systems. The system shall also be designed for extreme weather conditions; if the players are able to practice outdoor in Arctic conditions, the system should also be operational. The weather resistance and toughness of the players and coaches provide the limit, not the system.

Scalable: The system must scale along multiple axes. Video footprint persisted on disk should be small, which implies that only explicitly selected video segments are persistently stored.

Incremental growth of additional CCR cameras should be supported, as well as incremental growth of end-users. It is worthwhile noticing that different coaches can have different notational markers defined for the same practice depending on role and responsibility. As such, the human-computer interface should be easily, fast, and just-in-time configurable.

Economy: We are interested in developing a practical system that potentially scales well outside resource savvy sport organizations. The components in the system should primarily be cheap, off-the-shelf components; a minimalistic system configuration should potentially be purchased for less than 2,000 USD. This is orders of magnitude cheaper than some of the more popular notational analysis systems, not even accounting for avoidance of the costly human operational cost that competing systems carry. Also, Muithu should be designed for cloud computing integration so that additional expenses are expenditure related leveraging off cloud computing offerings.

Soft real-time: To scale into daily operational use, it is important to avoid tedious, human-centric post-processing of notations captured during a sports event. Ideally, the system should be operational in real-time, but for our application scenario, soccer coaches primarily need output from the system when the physical exercise is over.

Sport agnostic: The system should be relatively sport agnostic and easily applicable to almost any sport. In fact, as long as a real-life feedback scenario includes visual perception, it can be a potential application area for Muithu. This potentially includes application domains as e-learning and unified communication platforms.

III. RELEVANT GUIDING PRINCIPLES

The system architecture captures the logical structure of a system and guides developers in reasoning about, for instance, needed components and their interaction. This section details the fundamental design principles and one specific requirement that in particular constrained and influenced the system architecture and design.

A. Fundamental Design Principles

A set of fundamental principles and abstractions in computer system design is common across different applications and systems. While building a series of video-based notational systems, we applied, in particular, four such fundamental principles.

First, the well-known end-to-end principle in systems design [5] made us remove a large human-centric analytics component. In essence, this principle suggests that functions placed at low levels of a system may be redundant or of little value when compared with the cost of providing them at that low level. This provides a rational for moving certain functions upward in a layered system. Since operational notational systems often contain a human-centric analytic part at a lower level, we suggest moving this functionality as high as possible.

A closely related rationale for doing this is that analytics performed by others than the head coaches themselves might fail or be incomplete since only the expert coaches might have the capability of determining and value certain complex events. The net effect is that we literally speaking move notational functionality into the hands of the expert coaches.

There is a potential tension situation by adding technology distractions to the critical work path of the coaches, so this new technology must be non-invasive and retrofitted seamlessly into their working habit. If successful, though, the high recall and high precision requirements can be met. This also supports the soft real-time requirement since the need for human post-processing is virtually removed.

Our second design principle is to make use of upstream evaluation [6]. Consider a streaming substrate consisting of data producers and numerous, remote data consumers. A scaling technique is to provide expressive data filtering and correlation with potential consumer interests as close as possible to the data sources. This way, data not needed by remote consumers is filtered out and not shipped over the wire.

We apply this principle to meet the requirements of Muithu related to precision, soft-real time, and scalability. That is, we only transfer explicitly notated events from the video sources to higher layers in the system and leave the majority video footage rendered not needed at the sources. This might not sound as solving a practical problem, but end-to-end video transfer and conversion latency time from our CCR video cameras to a computer via an USB 2.0 interface is surprisingly high; the transfer and conversion time of a 10 second video of acceptable quality take approximately 20 seconds. Hence, we filter out and transfer only those short video sequences related directly to notated events. This approach leaves all redundant video footage at the source for garbage collection or later batch transferring.

On the path towards being persisted as unique, notational events, further filtering and reduction happens. For instance, input from non-relevant cameras or redundant events not really worth saving can be discarded. The idea is to avoid as much as possible of unnecessary transfer, processing, and storage to scale the system and to enable responses in soft real-time. For instance, even if an event is noted numerous times during a practice, a single video footage of the semantically same event might be sufficient for feedback purposes. This way we convert enormous amounts of data into structured, manageable information. To capture frequency if needed, a cost-efficient, footprint-greedy number can indicate how often the same specific event was noted.

A third and related design principle is downstream distribution [6]; data sent to multiple recipients should be disseminated as late as possible downwards a distribution stream. This avoids redundant communication, especially if the data source is a constrained resource. Our sources are in particular constrained by power and bandwidth limitations.

Finally, we treat the potential for incremental growth as a first order design principle. That is, we design the system so that it can scale incrementally upon demand without changing fundamental parts of it. The initial architecture and design of the system must be robust and extensible so that increased input sources, processing and storage demands, as well as new types of users and functionality can be accommodated. Agile and rapid scaling potential into cloud computing platforms, also known as elasticity, is an issue we will take into account where applicable.

B. Disruptive Potential

Explicit economy constraints are sometimes abstracted away by system architects. We cannot do this since there is a cost requirement associated with the Muithu system; it must be orders of magnitude cheaper than established computer and video-based notational analysis systems, and this combined with almost no operational cost involved. We target a digital management system as cheap and human-independent as possible, but with applicability potential way outside the limited group of resourceful sports clubs that use computerized notational analysis systems today.

The larger context for making the economy model explicit from the ground up is related to our search for disruptive technologies [7]. We intend to build a system that can accelerate disruption in the sport analytics and notational market, in particular providing a system for the masses. In Norway alone, approximately 2.112.000 of the 5.000.000 inhabitants are currently member of a sports club affiliated with the Norwegian Sports Association. Many of these sport clubs do not have technical competence or resources to acquire costly state-of-the art industrial systems, but could leverage off from our light-weight approach based on off-the shelf components.

A disruptive technology can informally be characterized as being considerably cheaper and have less functionality than existing state-of-the-art, but with the potential to radically improve existing or create new markets in unexpected manners. In the broader perspective, a low-end system with limited functionality entering a market may over time evolve into a role where the new technology expands and dominates the market.

To have disruptive impact, though, any such claimed novel technology must be paired with a business model the technology enables [8]. We leave the subject for now, but also leave the message that we actually intend to construct as cheap a system as possible with potential to open new use cases. This is in short a digital management system for the massive sport coaching domain, not only the elite part of it.

IV. The Muithu Architecture and Design

The Muithu architecture and design are directly impacted by two other high-end multimedia systems we previously have developed. One is a multimedia data collection, processing, storing, and streaming solution in the weather sector [9]. This system was probably the first academic multimedia sensor network ever deployed, and it is still in operational use after almost two decades of continuous operation and appearance on the Internet. A relevant lesson from this work is the befitting of a many-tier streaming architecture, where raw data is digested, filtered, and transformed into information while flowing towards end-users.

The other system directly impacting Muithu is a high-end multimedia search, video composition, and video streaming solution we initially developed for the soccer domain [10]. One deployment of this system is at the home field of Tromsø IL, where stationary CCR HD cameras cover the entire field. This

data can be manually analyzed similar to state-of-the art notational analytics systems.

Ongoing research investigates how to stitch together multiple camera views creating extreme wide-screen or 3D viewer experiences. We are also attempting to apply image analysis software to, for instance, detect specific events. To illustrate a surprising problem, we are currently striving with practical, real-time vision technology solutions detecting where the soccer ball is on the pitch.

New types of entertainment experiences are also developed, where viewers automatically subscribe to more detailed views of specific players or area of the field. Lessons learned include how CPU and I/O intense demanding this is, where a high-end computer is needed just to transcode and store input from maximum two CCT cameras in real-time. A typical footprint of such a camera is 500 Gbytes for a 90 minute soccer match. Experience from where computer, communication and labor resources needed to be spent thus impacted architecture and design details of Muithu.

Building Muithu has been an engineering enterprise where its architecture has been constrained and molded by the design principles outlined and the properties mentioned above. A naïve first stab at such an architecture might suggest a two- or three-tier layering scheme, but our architecture contains six main layers; the monitoring (1), the data collection (2), the raw data storage (3), the application (4), the information storage (5), and the end-user layer (6). This complies with our experiences from operating this type of multimedia video sensor networks for over two decades.

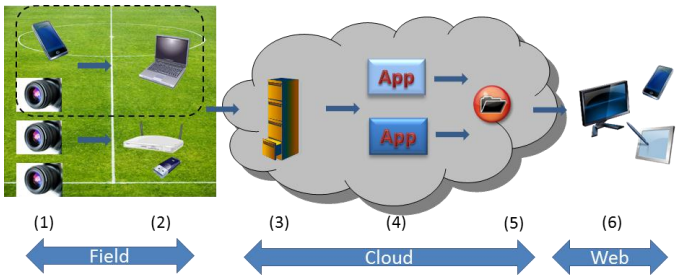

Figure 1. Muithu system architetcture.

Figure 1 illustrates the system architecture of Muithu. A typical deployment is spatially divided into three distinct groups: the components used on the sports field, the back-end modules designed for cloud deployment, and the web-based human-computer interaction components.

The monitoring layer (1) captures input during the actual sport happening. This can be a specific training exercise lasting minutes to complete 90 minutes soccer games. The minimalistic approach we are constrained by made us search for outdoor-resistant, fully automated, off-the-shelf video cameras. We selected a candidate motivated by experience from ski bums and glider pilots in our Arctic vicinity; the GoPro HD HERO2 Professional camera. This camera can be acquired for less than 500 USD, is weather and shock resistant, and captures professional full 170° (alternatively 90° or 127°) wide angle video footage at 1080p (alternatively 720p or 960p). The batteries has shown to last for approximately 2 hours in subzero conditions, so complete training exercises or games can be fully captured.

The mobile notational device candidates for the coaches included PCs, tablets, and cellular phones. We discarded the first alternative quickly due to weather-sensitivity, size and shape, but explored further the two latter by building complete user interfaces for each. The choice made by coaches moving about on the training field was clearly the least invasive and ubiquitous one, the cellular phone.

Hence, we selected a cellular running Windows 7.5 OS and implemented notational software in C# using Windows Phone SDK 7.1. A database containing the objects, in this case players, is kept with their respective events. These settings are easily configured, making it a minimal effort for users of the system to choose, add or delete objects and events.

We conjectured that a tiles-based interface would be convenient for capturing coach notations within our 5 second deadline requirement. Figure 2 illustrates snapshots of the interface. A notation is either captured by just clicking once at a player tile or by holding the thumb over a player tile and dragging this over personal evaluations goals appearing in less than half a second. The number of visible objects affects the size of each tile. As a consequence, it is possible to configure the application for coarse-grained or fine-grained annotations.

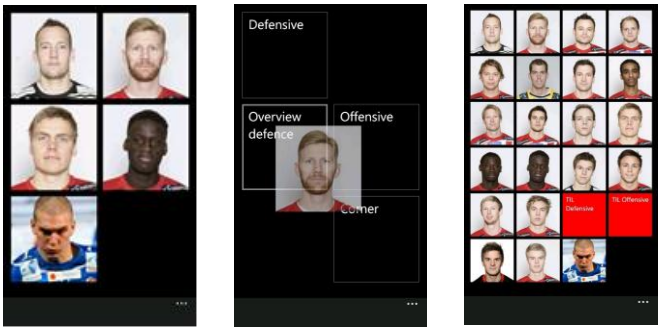

Figure 2. Tile-based interface.

Figure 3 shows the concrete sequence of notational data that this notation creates. This XML data will be stored and later used to fetch the corresponding videos from the active cameras.

```
<camsync id = "23">
        <camName>Overview north</camName>
        <timeStamp>28/04/2012 17:49:29</timeStamp>
</camsync>
<event id = "54">
        <objectName> Fredrik Bjork</objectName>
        <eventName>Defensive</eventName>
        <timeStamp>28/04/2012 17:58:03</timeStamp>
</event>
```

Figure 3. Notational data.

The notational process in Muithu is novel, but simple, once understood. Its simplicity though is the key to build a digital management system meeting the enlisted requirements. The essence is that the coach notifies an event *after* it has happened, which means that the coach observes the entire situation and determines afterwards whether it was a notable event worth capturing.

This *hindsight* notation approach relates directly to the end-to-end principle since the expert user evaluates and determines a posteriori whether the event was worth capturing or not. Only then is it persisted as a notation, which, simple as it sounds, provides high recall and precision. No other expert analysts or artificial intelligence software is needed.

Figure 4 illustrates Muithu notations plotted on a timeline. The Muithu client on the notational cellular phone is first started (S_{cell}). This starts the master clock all devices will be synchronized against. Whenever a camera starts recording, a start notation ($S_{cam1\text{-}n}$) is captured simultaneously to synchronize the video data stream with the master clock. When all cameras have been started and are synchronized with the notation device, the sport specific notations can be captured. Such notations ($N_{1\text{-}m}$) relates to the master clock, producing an accurate index into the videos continuously being recorded.

The head coaches have determined that 16 second video sequences are appropriate for game notations, with a start offset 15 seconds prior to the notation time combined with 1 additional second afterwards. In Figure 4, N_1 functions as a reference pointer into two video sequences, the same with N_2.

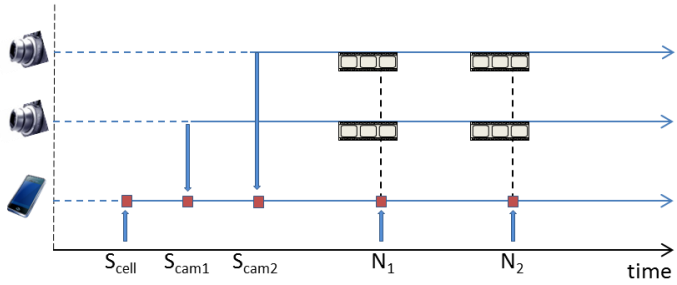

Figure 4. Notations captured

The data collection layer (2) in the architecture transfers data from the mobile devices to more persistent storage (3). The causality order among the devices dictates that the notations captured at the cellular phone have to be transferred first. This notational data functions as an index into the raw data still stored at the video cameras. The principle of upstream evaluation is implemented by simply pulling only the indexed sequences from the camera disks. The bulk data not explicitly requested will remain back at the camera as a rule and will later be garbage collected.

There are two ways to transfer data to persistent storage. First, the explicitly tagged video sequences can be synchronously transferred directly into the persistent storage (layer 3). So far, we can use USB 2.0 or insert the SD-card directly into the machine.

The second alternative is to add a proxy storage device, a mobile computer intended to be used at or near the practice field. The dashed rectangle to the left in Figure 1 illustrates the minimalistic configuration of such a Muithu system. This enables the coaches to cradle the cameras and notational data on premise, request the annotated video sequences, and view them for immediate evaluation and feedback purposes.

A side–effect, using the end-to-end and upstream evaluation principles this way, is that the expert coaches also can discard redundant, wrongly captured, or low quality sequences while on premise. For instance, if four cameras cover the same captured event, three of them might be discarded. Only the washed data is later transferred further into more persistent storage (3). Figure 5 provides an example of the human-computer interface coaches relate to while using Muithu on-premise.

Figure 5. Muithu on-premise interface.

The layers 3 to 5 in Figure 1 were built on a Microsoft enterprise cloud platform, but are currently being ported to a Microsoft Azure cloud computing platform. A REST API between the layers 2 - 3 and layers 5 - 6 simplifies this porting due to its loose coupling.

The persistent storage (3) is currently a Microsoft SQL database and a NTFS file system running on an enterprise cloud infrastructure (Intel Xeon W3550 3.06GHz CPU, Windows Server 2008 R2 Standard). The video sequences are stored as NTFS MPEG4 files, while the accompanying meta-data is organized in a traditional database table structure in a Microsoft SQL server. Requested videos are streamed remotely using HTML5.

Layer 4 is the application layer, currently built around Microsoft Internet Information Services (IIS) 7.5. One such application resembles the traditional one-to-many pedagogical experience between a coach and athletes. The coach typically prepares a presentation using video footage for illustration purposes, and next run through this orally with players present. Remote participants can join as well using, for instance, MS Office Live Meeting, and the whole session can be captured for later use.

A second and novel application is a one-to-one e-learning pedagogical application, which also is a closed enterprise social network. The players normally leave the training field for restitution and the like, but they can log into a web-based social

network from anywhere and engage in feedback dialogues using video sequences for illustrations. Enabled by technology can this interaction be asynchronous. A challenge-response dialogue has emerged as popular, where the coach can post a specific video sequence from the last practice together with a pedagogic question to a specific player. The player must next suggest alternative tactical moves or passes that better would resolve a specific problem documented in the video.

One motivation for this deeper and analytic treatment of the physical performance situations is to internalize coaching experiences in the athletes, not just tell them what to do. A peer-to-peer dialogue among multiple players can also be carried out.

A third application resembles state-of-the art analytics systems. Post-production of notations and expert analysis can be carried out on longer sequences that have been tagged with a start and stop button on the cellular. One extreme example is an entire game lasting 90 minutes, where an ontology-based soccer scheme can be used to manually notate events.

The coaches primarily wanted this capturing functionality for high-intensity sessions where all members of the squad participate. They can then view this video more in detail afterwards in entirety and select more events to be used in the pedagogic dialogue. For instance, Tromsø IL performs a very intense 16 minutes exercise once a week on a restricted field as an evaluation of the training details and focus the group has had the last week. The coaches can annotate as usual while this happens, but find it useful to also have the entire video footage for drill down analytics in more relaxed atmospheres.

Video search is also provided as an integrated application in Muithu. Coaches can, for instance, query and correlate players with specific events and get a storyboard presentation of related videos. This is a toolkit in particular used to capture temporal patterns and developments.

The information storage layer (5) contains specific video sequences and meta-data produced by the application layer. Each player has a folder containing links to videos (in layer 3), and related interactive dialogues from the past (from layer 4 applications). Privacy and security are implemented by password-authentication combined with associating a registered user with a set of roles to determine the authorization level.

The user layer (6) is primarily a web-based front-end. We are currently using Microsoft IE 9 due to its support for HTML5 and in-browser MPEG4 video viewing. Figure 5 illustrates a concrete interface snapshot from Muithu. Different cameras can be selected for viewing, deletion, further distribution to others and the like.

Surprisingly, the coaches often prefer overall views covering a larger area, not close-up scenes from individual situations. Of particular interests are team player patterns, relative spatial positions among individuals, and what those athletes far away from the ball are doing. Figure 6 is an example snapshot of this human-computer interaction.

Figure 6. Example interface Muithu.

V. EVALUATION

We have evaluated Muithu along two dimensions. The system performance and applied experiences are both based on empirical data from operational use.

A. System Performance Characteristics

The mode of operandi for transferring notational data and video into Muithu works as follows. First, the meta-data from the cellular phone is sent and stored on a computer, either through cable or WiFi using HTTP POST. The events describe timing events indicating when a specific camera is started or high-level notations. Timing events is used to synchronize and find offsets into videos at the different cameras relative to the real physical time as determined by the cellular phone master clock. Figure 4 illustrates the relative dependences of the different events plotted along a time axis.

A high-level description of the process from annotating to viewing an event is as follows:

Mobile

1. Upon notation input: store event name and current timestamp.

2. Upon push: HTTP POST of event formatted in XML.

Computer

1. Upon HTTP POST request: GET and parse the XML data.

2. Using the video start timestamp t_0, t_{event} and configured event length n, extract corresponding video from the source camera disk with an offset (t_{event} - t_0 - n).

3. Persist encoded video of the annotated event and meta-data.

Meta-data provides an index into the videos stored at the different GoPro cameras. We measured the transfer time of a 16 seconds video segment from a camera connected to a computer via a USB cable. The mean time observed for encoding and permanently storing a single event is 32.30s with a 95% confidence level of 0.44s. The numbers are calculated from 45 events on an Intel E6600 CPU with 4Gb of RAM.

The back-end store of Muithu receives videos, notations and other meta-data through a regular HTTP POST API. We have identified the pipeline steps involved upon receipt of a

compressed video message and conducted experiments to investigate their performance cost. The video sequence measured is equivalent to a 16 second sequence (3.711Mbytes), and we sampled 10 times. On average (confidence interval within 97%), end-to-end latency from receipt of the message and until the video is stored on local disk and the meta-data in the database is 0.784 seconds. The average cost of the individual steps is depicted in Figure 7. Notably, the processing of thumbnails accounts for the majority of the total cost.

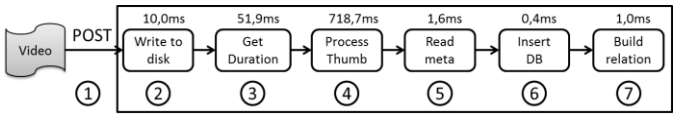

Figure 7. Processing steps of back-end storage server.

Another interesting performance indicator is the disk footprint cost. 90 minutes compressed video in acceptable format (MPEG4 2048kbps) from a single GoPro camera averages 1.336 Gbytes. The disk footprint of 20 sequences, each 16 seconds, accumulates to 74.22Mbytes. The reduction factor and disk cost saving is thus significant; disk footprint of 20 video sequences constitutes just 5.56 % of the total disk footprint of the entire game. This complies with one of our scale requirements.

We explicitly stated as a property that notations should be captured with minimalistic and non-intrusive efforts. The acceptable average event notation process was defined to be less than 5 seconds. By carefully observing the notation process during operation, we anecdotally observe that notation duration can be clustered around three duration sequences. The most frequent duration is when the coach is holding the mobile device during operation, and a notation then takes in the order of 2-3 seconds. Figure 8 shows coach Christensen at the reserve and staff coach bench using Muithu this way during a recent Premier League game.

Figure 8. Notation captured during game.

The second cluster is when the device is placed temporarily in a pocket for a while giving notation durations lasting 4-8 seconds. The third time cluster is more considered exceptions. Some problems taking notations were observed while being used in freezing sleet during late Arctic winter conditions; a wet cellular that needed to be wiped with soggy gloves added considerable time. Though, in general, we consider notation time as non-invasive.

We also conducted a more controlled and detailed study of the notational process by capturing Christensen on video during the last game in the spring season 2012. By using the notations from the cellular, we extracted the corresponding videos of the coach and manually clocked the actual time used for notational purposes. We defined time to be from when he shifted eye focus from the game, the time spent looking at the cellular and performing notational input, and until his focus was back on the play again. The mean time used is observed to be 3.1 seconds. Figure 9 depicts the distribution of notational time for this specific game.

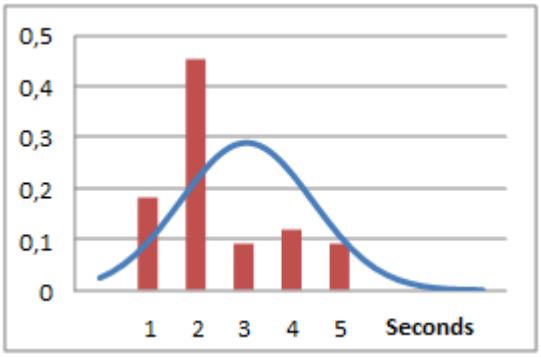

Figure 9. Distribution of notational process time (27. May 2012).

B. Applied Experience Characteristics

The crux of the Muithu scalability and real-time performance is the simple, yet not so intuitive concept of hindsight recording. The idea is that a user marks the end of an event worth capturing, not the beginning of it. The system will later find the beginning of this event, typically preconfigured to be 15 seconds prior to the end notation and one additional second. The conjecture we have is that this will dramatically reduce the number of notational events captured in real-time.

When used initially, we explained this principle to head coaches. To our surprise, we anecdotally observed that the coaches used Muithu as a more traditional recording device; they actually noted the beginning time of what potentially could materialize into a notation worth capturing. The very first notational rate was therefore in the order of one event every 23 seconds, and the coaches admitted that they actually had problems logically with hindsight recording. We attempted to explain this concept more thoroughly prior to next practice, and the frequency of notational marking went down to approximately one every 103 seconds.

Further training, in particular by sitting down with the head coaches and visual storyboards with all their notations and accompanying videos, was educational. The notation rate dropped rapidly over the next month in operation, and we soon

observed that the frequencies during training sessions dropped closer to one event noted every 4 minutes or so.

We then conducted a more thorough analysis to get less anecdotic numbers. Hughes et al. [11] suggested 6 soccer matches as representative for defining performance profiles, so we used a total of six workloads gathered from real matches against teams at the same level as Tromsø IL. Figure 10 depicts the number of notations performed by one of the head-coaches during 6 consecutive games spring 2012. It averages around 16 notations per match, but with two notable exceptions.

First, the match against Hønefoss produced 28 notations, most of negative nature. This was a game ending with a draw 0-0, but where Tromsø IL clearly under-performed. Second, the game against Viking produced only 5 notations in a game Tromsø IL controlled entirely and comfortably won 5-1.

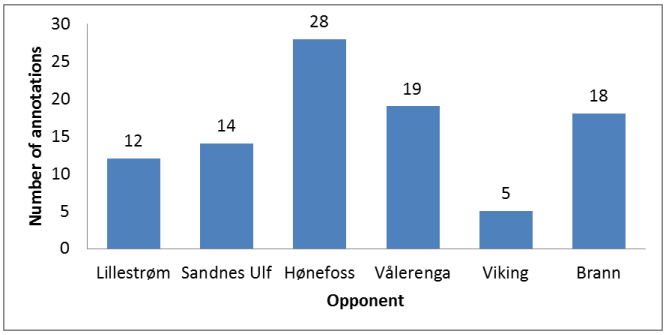

Figure 10. Notations captured during games.

An interesting comparison is with other notational analysis systems. Tromsø IL also uses ZXY Sport Tracking [12], a radio based monitoring system capturing performance measurements of the individual athletes. This system operates with a frequency of 40 samples per second, so just the position of a single player accumulates 216,000 notations during a 90 minutes match. With position, direction and the like, over a million samples is noted per player per match.

We defined recall R as the fraction of the important events triggering a coach response during a session that are successfully recollected. As such, we have close to perfect R = 1 in all workloads.

One exception experienced was during the last of our 6 games used in our experiments. This game was carried out in sub-zero weather conditions, with snow in the air and sleet on the ground. The coach kept the soggy cellular in his pocket most of the game, but we experienced a sudden spike in number of notations taken. Normal notational frequency is in the order of less than 20, this game had 431 notations.

Since we have a video camera explicitly recording the coaches during games, we could tediously select all 431 video sequences and manually examine when the coach actually performed a notational action. We could then determine through visual perception what was really a notation being taken, and what was false positives (from a frozen hand in the pocket touching the screen). Obviously, the cellular in a soggy pocket and an engaged coach had resulted in too much input as illustrated in Figure 11. On a side-note, the correct number of real notations for this game went from 431 to 18.

```
<event id = "128">
        <objectName>TIL Offensive</objectName>
        <timeStamp>28-Apr-12 6:25:54 PM</timeStamp>
</event>
<event id = "129">
        <objectName>TIL Offensive</objectName>
        <timeStamp>28-Apr-12 6:25:54 PM</timeStamp>
</event>
<event id = "130">
        <objectName>TIL Offensive</objectName>
        <timeStamp>28-Apr-12 6:25:55 PM</timeStamp>
</event>
<event id = "131">
        <objectName>TIL Offensive</objectName>
        <timeStamp>28-Apr-12 6:25:56 PM</timeStamp>
</event>
```

Figure 11. Too much notational input.

The other interesting parameter we are interested in is precision P defined as the fraction of retrieved instances that are relevant. Intuitively, this is one, even if the coaches admit that not all captured events have the same value and validity in a longer perspective, for instance displayed to all players in the social network.

We need to do a more careful and extended study of precision in a data life cycle perspective. We are, for instance, currently developing trend analysis applications correlating incidents and recurring patterns over several games.

To do this coaches have started to explicitly annotate the most valuable events and camera angels from each game. To illustrate precision in this context, we have plotted the total videos captured versus those finally persisted in this context for the two first games in our series. As can be observed in Figure 12, there are even less video sequences worth keeping in this longer perspective.

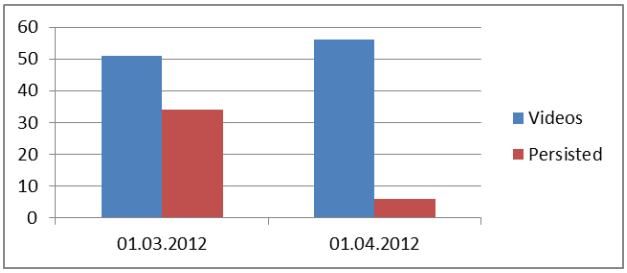

Figure 12. Total videos captured vs persisted.

VI. RELATED WORK

Different technologies for automatically capturing athlete performances have been put in operation. Interplay-sports [2] is a system being used since 1994 where video-streams are manually analyzed and annotated using a soccer ontology classification scheme. Trained and soccer-skilled operators tag who has the ball, passed carried out and the like. The more skilled the operators, the more soccer specific events can be

derived. Yet, we consider this type of system to primarily provide statistics, not necessarily advanced analytics.

This process can be automated. Computer vision algorithms provide the foundation for tracking multiple athletes in a video stream and can be used to determine, for instance, activity profiles of athletes and examine high-intensity running [13]. Visual tracking of multiple targets is a challenge, but solutions have been proposed for multi-people tracking that preserves identity of athletes even while temporarily hidden by others [14].

ProZone [3] is a commonly used commercial system using video-analysis algorithms. In particular, it quantifies lower-level movement patterns and characteristics like, for instance, speed, velocity, and position of the athletes. Di Salvo et al [15] conducted an empirical evaluation of deployed ProZone systems at Old Trafford in Manchester and Reebook Stadium in Bolton. They concluded that these 8 CCR video camera deployments represent accurate and valid motion analysis.

Existing video-based analysis systems, surprisingly as it might sound, have a problem with the soccer ball. State-of-the art vision technologies cannot successfully and practically identify and capture a regular soccer ball on the pitch with acceptable accuracy. We consider this as one important reason for why systems like ProZone still need a considerable human component in the analytics process. The other reason is that state-of-the-art artificial intelligence solutions still cannot render the human expert superfluous.

Global positioning and radio based systems are alternative technologies for capturing performance measurements of athletes. A main difference with video based systems is that monitoring and transmitting equipment must be carried by the athletes. ZXY Sport Tracking [12] is one such system that provides detailed physical athlete information like, for instance, player speed, orientation on the field, position, step frequency, and heart rate frequency with a resolution of samples 40 times per second.

We consider physical measurements of individual athletes as low-level notations. Human expert annotators extracting soccer information from videos are either medium- or high-level notations. The difference is whether the notations process is carried out by the head coaches or somebody else. Experience is that head coaches as a rule have outsourced this notation task due to time-constraints. Hence, systems like ProZone then provide medium-level notations.

Muithu provides high-level notations. This is not because we consider Muithu superior to the other technologies, but because it captures real annotations by the highest level experts, the coaches in the midst of the situation. Muithu is thus primarily complementary to the afford-mentioned technologies.

As anecdotal evidence supporting this claim, Tromsø IL is in fact using low-, medium-, and high-level notational systems. ZXY Sport Tracking provides detailed physical and geo-location information, Interplay-sports provides medium-level analysis done by additional experts, while Muithu provides the high-level notations by the head-coaches themselves.

VII. CONCLUDING REMARKS

A rapidly growing notational sport analytics market is spurring companies providing systems that capture speed, velocity, position and the like of sport athletes. Unfortunately, all these digital information management systems still depend on human experts post-analyzing and annotating more complex sport events. This is where our non-invasive in-game hindsight recording concept provides a new and novel approach. It captures incidents by expert coaches that potentially determine successful or failed performance based on the concept of hindsight recording. This is the fundamental concept that enables a combination of a diverse and orthogonal set of requirements in practice. Since coaches determine a posteriori whether an event is worth noted we are applying the end-to-end principle in practice.

Despite technology being massively introduced throughout society, sport coaches surprisingly still use pen and pencil as a rule for capturing performance related events in real-time. This is typically followed by oral feedback to the athletes. We conjecture that this lossy, one-way feedback process is ripe for change as we have demonstrated a proof-of-concept implementation of a mobile, light-weight, non-invasive notational analysis system that is close to fully automated. It provides highest recall possible, but also automated correlation and provisioning of the related video segments published through an enterprise social network. Coaches and athletes now have a novel multimedia toolkit for interactive feedback, peer-to-peer discussions, and e-learning purposes.

Muithu has also demonstrated proof-of-performance by actually being able to run and scale in practice for a Norwegian Premier League soccer club. We also consider Muithu as a proof-of-applicability system since it has been evaluated and rendered useful by expert users, least said that it has been put to operational use.

It has been a new, but interesting and compelling experience for us computer scientists to constrain our design and implementation alternatives while developing Muithu. Our mode of operandi is normally to improve on high-end cutting-edge research technology solutions, now we had to limit our solutions to low budget constraints. At the same time, we designed Muithu so that it can leverage off cost-efficient cloud scaling potential whenever more resources are needed. We consider our carefully carved, cheap, low-end solution to have similar novelty and impact potential as state-of-the art large footprint systems.

ACKNOWLEDGMENT

This work is funded as a Norwegian Research Council Centre for Research-based Innovation over 8 years.

REFERENCES

[1] http://www.btlnet.co.uk/media/1357546/btl-cisco.pdf

[2] http://www.interplay-sports.com/

[3] http://www.prozonesports.com/index.html

[4] Franks, I.M. and Miller, G., Eyewitness testimony in sport. Journal of Sport Behavior, pp. 39-45, 1986.

[5] Saltzer, J. H., Reed, D. P, and Clark, D. D., End-to-End Arguments in System Design. In: Proceedings of the Second International Conference on Distributed Computing Systems. Paris, France. April 8–10, 1981. IEEE Computer Society, pp. 509-512.

[6] Carzaniga, A., Rosenblum, D.S., and Wolf, A.L., Design and Evaluation of a Wide-Area Event Notification Service. ACM Transactions on Computer Systems, Vol. 19, No. 3, August 2001, Pages 332–383.

[7] Christensen, C. M., The innovator's dilemma: when new technologies cause great firms to fail. Boston, Massachusetts, USA, Harvard Business School Press, 1997, ISBN 978-0-87584-585-2.

[8] Christensen, C. M. and Raynor, M. E., The innovator's solution: creating and sustaining successful growth. Boston, Massachusetts, USA, Harvard Business School Press, 2003, ISBN 978-1-57851-852-4

[9] *Anonymous – to comply with ICDIM 2012 review constraints.*

[10] *Anonymous – to comply with ICDIM 2012 review constraints.*

[11] Hughes M., Evans S. and Wells J., Establishing normative profiles in performance analysis. International Journal of Performance Analysis in Sport, Volume 1, Number 1, 1 July 2001 , pp. 1-26.

[12] http://www.zxy.no/index.html

[13] Bradley, P.S., Sheldon, W., Wooster, B., Olsen, P., Boanas, P., and Krustrup, P., High-intensity running in English FA Premier League soccer matches. Journal of Sports Sciences, Volume 27, Issue 2, pages 159-168, 2009.

[14] H. Ben Shitrit, J. Berclaz, F. Fleuret and P. Fua, Tracking Multiple people under Global Appearance Constraints. IEEE International Conference on Computer Vision, November 2011.

[15] Di Salvo, V., Collins, A., McNeill, B., and Cardinale, M. Validation of Prozone: A new video-based performance analysis system. International Journal of Performance Analysis in Sport, Volume 6, Number 1, June 2006 , pp. 108-119.