Abstract—In this paper we describe how content-based analysis techniques can be used to provide much greater functionality to the users of an interactive TV (iTV) device. We describe several content-based multimedia analysis techniques and show some of these can be exploited in the iTV domain, resulting in the provision of a set of powerful functions for iTV users. To validate our ideas, we introduce an iTV application we developed which incorporates some of these techniques into a simple set of user features, in order to demonstrate the usefulness of content-based techniques for iTV. The contribution of this paper is not to provide an in-depth discussion on each of the individual content-based techniques, but rather to show how many of these powerful technologies can be incorporated into an interactive TV system.

Keywords—Interactive TV, content-based analysis.

I. INTRODUCTION

The history of iTV stems back to the early days of television itself in the 1920’s [5], and since that time there have been numerous attempts to launch iTV into mainstream use. Many of these have failed, and this may be one of the reasons why interactive TV is still often thought of as a new technology. With recent developments such as mobile TV viewing, the use of TV in conjunction with other interactive devices, and the Web 2.0 phenomenon for socially-oriented interactivity, iTV research today has many opportunities to move forward. However, there is another important technology that has yet to be incorporated into the iTV domain – content-based multimedia analysis – and we believe that such technologies provide exciting opportunities for enhancing the viewer experience.

One of the drawbacks to current iTV is that it can require large production efforts in order to manually create interactivity for TV content. As a more efficient alternative, we propose a number of content-based analysis techniques to automatically create interactivity – not only between TV data (provided by broadcasters), but also between the user’s own uploaded content, and later in this paper we show this working in practice.

The areas of computer vision, content-based analysis and multimedia information retrieval are established areas of research, however “despite the considerable progress of academic research in multimedia information retrieval, there has been relatively little impact of MIR research on commercial applications” [8]. This may not remain the case for long as research groups are beginning to investigate applying content analysis in consumer electronics [1]. Having developed a prototype iTV system that utilises these technologies we can see their benefit for this kind of system, and in this paper we describe some of the key features that these technologies can provide for iTV.

In Section II we give a brief overview of content-based multimedia analysis. Section III describes how content-based analysis techniques can be used to enhance the functionality of an iTV, giving specific examples of how we have integrated these into our own prototype system. Section IV describes our prototype system and finally in Section V we summarise the functionality that is provided from the integration of content-based analysis techniques, and we examine the impact that these can have on the iTV domain.

II. CONTENT-BASED ANALYSIS

Content-based multimedia analysis can refer to a variety of approaches which examine content itself directly, rather than only the meta-data associated with content [16], [8], [4]. For image and video data in an iTV context, content-based analysis analyses the “content”, which may include “low-level” (colours, textures, shapes, etc.), “mid-level” (face detection, etc.) and “high-level” features (identifying the semantic meaning of the image). Without the use of content-based analysis, a system’s knowledge about a video is limited to the meta-data associated with it. Such meta-data may be quite limited if automatically acquired (from an EPG for example), or it can be time-consuming and expensive to generate if created by a human. However, with the use of content analysis, a system can automatically generate a description of the video content, which will allow greater functionality for the end-user.

While not all content analysis techniques are effective or robust enough to be used in commercial applications yet (many of these are still at the research stage in laboratories),
we are starting to see early applications, e.g. a photo
organiser with automatic face detection [2], and we expect
to see more applications appearing in the near future. In all
we believe there are a number of very useful content-based
technologies that can be useful in enhancing the functional-
ity of an iTV device, including keyframe extraction; shot and
scene boundary detection; image similarity techniques; face
detection and face recognition; video summarisation; news
story segmentation; and text-based searching. Although it
is beyond the scope of this paper to discuss each of the
core content-based technologies in depth, we give a general
overview of each of the technologies being used and we
describe how we have utilised these, in order to provide
enhanced functionality in an iTV system.

III. Enhancing iTV Functionality

In order to demonstrate the usefulness of content-based
analysis and search techniques for the iTV domain, we give
specific examples of enhanced functionality which we have
integrated into an iTV, and for each of these we describe
the underlying technologies necessary in order to implement
these.

A. Shot and Scene Browsing

We are all familiar with the chapter-level browsing of
films that is facilitated by most DVDs. The creation of
these chapter menus generally requires manual intervention
and so can be time-consuming and expensive to generate.
However, with the use of content-based techniques, the
information for this type of browsing can be automatically
generated, thereby providing a browsing mechanism for all
TV programs that a user records on their own personal iTV
without the need for manual intervention.

One of the content-based technologies that can be used
to make this possible is shot-boundary detection, which
in the content-based analysis community, is considered “a
fundamental step for the organization of large video data”
[10]. A video or camera “shot” may be considered as
the basic unit of a video – shots are usually quite short
in duration and a video typically consists of many shots,
which are separated by a transition (e.g. cut, fade, dissolve,
wipe). Shot-boundary detection attempts to identify these
shot boundaries and allows the video to be broken up into
its constituent shots. In order to do this the image frames
from the video are firstly extracted, and then the detection
works by comparing adjacent or nearby frames in order
to identify where a shot-boundary occurs. Shot-boundary
detection is at a quite advanced stage in terms of research,
and although further research is being carried out in this
area, many consider this to be a “solved problem”.

Although video shots are useful for organising video, for
the user/TV viewer they (by themselves) are generally not of
much use. However, if these shots can then be aggregated to
form a more semantically meaningful unit of video, such as
a scene, then this is much more useful, and in fact using an
intelligent shot aggregation technique we can automatically
generate scene-level access to a piece of recorded video.
Once these scenes have been identified, we can use a
keyframe extraction technique to identify the most representa-
tive frame for the scene. A series of these keyframes can
be miniaturised and displayed to allow efficient browsing of
video content without having to play sequentially. Figure 1
shows a screen shot of our iTV application, demonstrating
the result from this technique. This provides scene-level
access for each video, as well as allowing browsing within
each video scene via the shots. Alternatively this could
be presented in a conventional system such as that used
in a DVD chapter menu for instance. Also if the iTV
device can support the user inputting their own home movie
footage for example, then the same approach can also be
applied to divide the user’s own video into meaningful units
and provide an easy way to navigate through their video
collection.

Related to this, there is research in the area of scene
classification which attempts to identify scene types e.g.
action, dialogue, montage, etc. (usually within movies) [7].
Being able to identify scene types that are present in a
movie can provide numerous possibilities to an iTV device,
in allowing users to interact with their movie content in
different ways. For example, the system may provide a
means by which to browse all the movie scenes recorded
on a user’s iTV based on the scene type, or while a user
is watching a scene they may want to find other similar
scenes (from within that movie, or from others) which are
of the same type. In terms of content analysis, there has
been a significant amount of research devoted to this issue
of identifying the type of scene. However at its current stage
this technology has only been successful in identifying a
limited number of scene types or classes, and so we have
chosen not to integrate this feature into our system, though as
this technology develops we envisage this feature providing
a useful method of browsing and linking through content.
B. Automatic Linking to Related Content

The facility to automatically link video to “related content” can enhance the interactive nature of viewing TV content. A user may wish to link to related content from within a program or film, for example wanting to see content that is similar to what is on screen (maybe because of the actor on screen, or because of the type of the content), or a user may wish to find something similar to an entire program or film – related to the genre or even simply the title of the program. All this can be automatically achieved through the use of content-based technologies (described in [3]). As linking between related content can be done automatically and can take into account the viewing context of the user, the user can easily find related content, without the need to input text into the system (which may be inconvenient to do when interacting with a simple TV remote control). Essentially the user can search, without the need to input text, by taking into account the context of what the user is watching on screen at the time of the search request, and the system can automatically find relevant information. This context can use both meta-data (automatically attained from the EPG) associated with the current show, as well as low-level content based features (such as colour histograms) in order to match similar content.

The same type of functionality may also be achieved using large amounts of user annotation of video, but as we know this can be time-consuming to generate. As content-based techniques automatically create this type of linkage they provide a more efficient and scalable alternative. There are a number of these techniques that can be applied, which we now discuss.

1) Text-Based Similarity: Text-based similarity provides a simple and straightforward method of finding similar content, allowing matching between TV programs based on any text associated with the program. One simple way of providing this matching using a program’s text (either the text within the program’s meta-data, or text within the subtitles) is to use a tf-idf approach [14], which is a classic text search strategy. This should be quite effective in finding related content, particularly when searching relatively short sources of text – such as the program title and description of generation of movie trailers [9]. We have found that for a video, without having to watch the actual content.

D. Program Previews

The ability to display a preview of a recorded TV program is a useful facility, particularly when dealing with browsing a very large collection of recorded content. This can provide a user with a quick overview of what is contained within the video, without having to watch the actual content.

There are a number of techniques that content analysis can provide in order to generate a preview video for this purpose. As previously discussed in Section III-A a video can be broken up into a number of shots, and if we can provide an identifying keyframe (or number of keyframes) from each of these shots, and then play back only these keyframes, we create a sequence that in itself can serve as a simple preview of the video content.

Related to this, there are many different approaches for generating summaries of different genres of video content, and in particular a lot of research has focussed on the area of generation of movie trailers [9]. We have found that for most popular movies it is possible to automatically acquire a user-created movie trailer on the WWW and so our chief concern has been the summarisation of content for which we cannot readily retrieve an existing summary from an external resource.
E. Sports Event Detection and Summarisation

Specific content, such as sports, has certain characteristics that make it quite different from general content, and so we process sports content differently from other content, in order to provide a more meaningful summary.

In recent years sports event detection and summarisation has received quite a bit of attention by using content-based analysis [17]. This research has developed techniques for automatically detecting important events in sports programs, and in using these to generate a summary video [13]. The ability to record a soccer match (for example) and then later to have the option to view only the goals, or a quick 5 or 10 minute summary of the game seems like an extremely useful feature to provide to an iTV user.

Our approach to this is to firstly identify a set of key events for each sport that we wish to work with (for example a goal is a key event in soccer). We then identify a number of key features that we can work with in order to determine when a goal has been scored, e.g. crowd noise, player close-up, etc. We then use a machine learning approach in order to learn these key events and to build an automatic classifier [13]. Each sport uses a separate classifier, as each sport has different indications of important events and so each of the features should be combined differently – in our system we provide sports event detection on soccer, baseball, rugby and gaelic football.

The way we integrated this into our iTV system was, to allow more effective browsing of the sports broadcast by providing access to the most important events in the game, in a similar way to allowing access to scenes (as described in Section III-A).

In addition to providing this timeline-based means of navigating through the events in the sports show, we have also used the underlying content-based sports event detection tools to predict the most important events within the game. This allows us to provide an alternative means of viewing the sports show, by giving a visual representation of the entire show on screen, as illustrated in Figure 2. Here the most important events within the game are represented as larger keyframes on screen – allowing the user to quickly navigate to the key events in the game. Also, as the key events within the sports show can be identified, it is possible to generate automatic sports highlights for each of the recorded sports shows.

F. News Story Segmentation

News content is also quite different from other content, and in addition to treating it differently, to provide more meaningful linking to related content (as discussed in Section III-B), we also treat it differently when generating a means to browse through an entire news broadcast, as well as specific news summaries.

The ability to skip through a news broadcast to a specific story of interest is also a highly desirable feature for iTV. Although this type of functionality can be provided with the use of manual annotation, which can be expensive and time consuming, in order to do this automatically we can employ the use of content analysis techniques. Similar to sports event detection, news story segmentation has also received significant attention in the field of content analysis in recent years [15]. In our work on iTV there are a number of approaches that we use in order to segment an entire news broadcast into its individual news stories [12]. The identification of “anchor-person” and other types of “within studio shots” in a video broadcast, is usually a strong indication of a news story, as well as the appearance of certain logos (for particular news channels). These can help us to identify a news story boundary and segment the news into individual stories. Also for certain broadcasts that complement the broadcast with subtitle text, it may also be possible to analyse this text in order to help with the news story segmentation.

Once news story boundaries have been identified we can provide a mechanism on-screen for the user to easily browse between stories. Our implementation of this is essentially the same as that provided by the scene browsing function (as described in Section III-A) except that in this case each thumbnail represents a story in the broadcast, and selecting any of these will cause the video to jump to the start of that news story. So although from the user’s perspective, the way in which they browse through news content will be similar to that of general content, the points that they browse to will be much more meaningful as they correspond to story boundaries – which allows a user to jump to a news story that interests them, and also jump to a related news story from a different news broadcast (as discussed in Section III-B).

IV. iTV System

We have developed an interactive TV system which presently runs on a desktop PC running Windows XP and is connected to a high definition TV. Currently interaction is carried out via remote control, with dedicated buttons to
launch the most common features in the system, as well as the standard up, down, left right and select buttons. All interactions are made via these buttons only, in order to mimic the interactions of a TV and to allow simple navigation for the user [6].

The following enhanced features (as discussed in Section III) have been implemented in our system:

The shot and scene browsing feature (Section III-A) has been implemented to allow easy browsing of any video content that is recorded and stored. We also take into account the type of video content in order to segment content differently (using underlying content-based analysis techniques), so for news content the system applies news-story segmentation allowing news story browsing [11]; for sports programs the system allows browsing of the key events within the game, as determined by our sports event detection [13] (discussed in Section III-E).

As part of our video archiving browsing (where the user can browse through their recorded contents), we use our program preview facility (Section III-D) to show a quick preview of the content, to help the user decide if they want to view the entire program.

Our system also provides a find similar feature, finding similar content to what the user is currently watching. The system does this in a number of different ways, using the approaches discussed in Section III-B, including using text meta-data to find similar TV programs from the user’s collection, as well as related content from YouTube and finding content which is visually similar to the content that the user is currently watching on-screen. This type of intelligent searching provides the user with the facility to do advanced searching, with the system taking care of all the complexities, using meta-data associated with the program, as well as the type of content that the user is viewing in order to provide the most appropriate similar content.

V. CONCLUSIONS

In this paper we outlined the benefit of including multimedia content analysis technology on an iTV. We discussed a number of key functionalities for an iTV, as well as outlining the underlying content-analysis technology we used in our implementation. We believe that the integration of these (as well as other) content analysis techniques can allow for much greater interactivity between different types of media, web pages, video, image, music, etc. Although in introducing these powerful new tools there is the potential to overwhelm the user with an array of complex functions and interactions – as well as the underlying technologies – we feel it is just as important to also take into consideration the usability of the system, and in our recent work we address how this can be achieved [6].

For our future work we plan to examine and evaluate these powerful technologies into an iTV device.

We also plan to evaluate the usability of the system through user testing in order to determine the best way of integrating these powerful technologies into an iTV device.

REFERENCES


