

Enrichment of Interactive TV Services with Collaborative and Content-based Filtering Methods

Peter Dunker
dkr@idmt.fraunhofer.de

Christian Dittmar
dmr@idmt.fraunhofer.de

Fraunhofer Institute for Digital Media Technology
Metadata Department
Ehrenbergstraße 31
Ilmenau, Germany

ABSTRACT

One of the most successful interactive TV applications are the mobile messaging TV services e.g. SMS chats. Within this paper we present novel approaches for enrichment of mobile messaging services utilizing content-based and collaborative filtering methods. We depict the current work of a content and playout server used for integrating interactive TV services in digital TV networks. Furthermore, we sketch developed content-based analysis algorithms which are able to enrich the user experience of mobile messaging services as well as offering opportunities for novel services with less service production costs. The new approaches concentrate on automated multimedia content acquirement and intelligent content scheduling and preparation.

Categories and Subject Descriptors

H.5.1 [Information interfaces and presentation]: Multimedia Information Systems; H.3.1 [Information Storage and Retrieval]: Context Analysis and Indexing, Abstracting methods, Indexing methods; D.2.11 [Software Engineering]: Software Architectures

General Terms

Algorithms, Design, Human Factors, Reliability

Keywords

Entertainment, interactive, mobile messaging TV, multimedia content-based analysis, collaborative filtering

1. INTRODUCTION

Currently there is a broad range of interactive TV services available in the world. The kind of applications can be categorized with different taxonomies: local individual interactive without return channel (information services) vs. full individual interactive with return channel (home shopping)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 200X ACM X-XXXXX-XX-X/XX/XX ...\$5.00.

vs. mass interactive (voting applications). An alternative categorization is: TV broadcast attendant (quiz applications parallel to TV quiz shows) vs. TV broadcast independent (TV operator service portals) vs. TV broadcast generating (mobile messaging chat applications).

In this paper we concentrate on mobile messaging interactive broadcast generating services such as the Norway TV show "Mess TV" [1] which incorporate Short Message Service (SMS) and Multimedia Messaging Service (MMS) messages send via mobile phones directly to everyone's TV screen and to talk with moderators. Furthermore, MMS messages are shown on the screen to more involve the viewers in the TV show. Various interactive TV broadcasts based on SMS and MMS communication can be found in [2, 1].

The purpose of this paper is to show the ongoing work of novel enrichment strategies of mobile messaging interactive TV services with collaborative and content-based filtering methods. Furthermore, we depict enrichment strategies and example service concepts with the focused commonality to need less production and maintenance support as possible for TV broadcast generation.

2. COLLABORATIVE AND CONTENT-BASED FILTERING

2.1 User-Generated Data Sources and Collaborative Filtering

User-generated data sources can be divided into two groups. First, the multimedia data that are provided by the viewers to control the current broadcast. These information can be images, text messages or voice message submitted via mobile phones or alternative available return channels. The multimedia items can be utilized directly to be part of the TV broadcast or can be used as seed media items within a below described similarity search process. The second group of data sources means incorporating user-generated data from Internet platforms such as Flickr, Wikipedia or blip.tv which offers a wide range of public available multimedia informations. In our use cases we concentrate on multimedia data that are published under creative commons license¹.

Collaborative filtering methods [3] can be applied to improve TV content recommendations by joining multiple user interactive feedback. Therefore, users of the interactive TV services need to be monitored e.g. with their kind of input to TV services which can be used to setup a user profile.

¹<http://creativecommons.org>

The profiles of users involved in a current interactive TV service can be compared to filter the most relevant.

2.2 Image Analysis

In the domain of image analysis two principle approaches are relevant for the presented work. First the content-based image retrieval (CBIR) which uses image examples as query to an image database and return similar images based on low-level features such as color histograms or spatial frequency information. Reviews of common techniques can be found in [4]. A typical result of our CBIR algorithm is depicted in figure 1. The second approach is based on classification of image content. This approach summarize different pattern recognition and classification techniques e.g. natural scene classification [5] or object recognition [6]. The results consist of semantic descriptions of the whole image e.g. landscape, city, people or descriptions of parts of the image e.g. grass, forest, beach, sea or specific annotations of concrete objects e.g. car, horse or Eiffel Tower.



Figure 1: Example results of our CBIR algorithm with visual similar images.

2.3 Audio Analysis

Semantic audio analysis refers to the enrichment of digitized audio recordings with descriptive metadata tags allowing for advanced database search or further processing. Nowadays, metadata description of audio is mostly given by catalogue-oriented classification. Unfortunately, this labeling according to predefined categories is cumbersome and time-consuming. Automatic content-based metadata generation promises cost-efficient, scalable mark-up, and the state-of-the-art is advanced enough to deliver robust and efficient results for real-world applications. In terms of audio retrieval by keyword-search the most interesting properties of a piece of music are its genre and mood. The most widely acknowledged musical styles and mood characteristics determined in user-studies can be classified automatically using methods of data mining and statistics in a supervised pattern-recognition system [7, 8]. Other useful properties can be robustly determined, such as the tempo, bar-type, and beat-grid. We exploit our temporal information analysis algorithm [9] for the automatic generation of image-sequences according to accompanying music. Furthermore, song structure segmentation [10] poses an important pre-requisite to account for changes of the genre or mood throughout a song.



Figure 2: Result of our song structure segmentation algorithm for the song "Madonna - Like a Prayer". Segments labeled with *C* identified as chorus.

An efficient way to schedule additional music content is content-based music similarity retrieval in case no semantic tags are available. A description of our approach can be found in [11, 12]. The computation of similarity relations

between a query song and a larger catalogue of songs can deliver relevant playlists. While extending our system to adapt such playlists by modeling user preferences and incorporating relevance feedback in the processing chain we could improve the recommendation quality [13].

The application of content-based music identification via mobile phone voice call holding the mobile phone close to a loudspeaker [14] offers further interaction possibilities.

2.4 Multi-Modal Analysis

A multi-modal or cross-modal analysis process includes audio and visual media. A typical multi-modal analysis application is a music player which uses photos for visualizations. The combining aspect of the the modality could be the music and image mood or emotion impacts. Our contribution in the field of multi-modal mood classification can be found in [15]. Chen et al. [16] evaluates their music visualization with user tests to measure the user perception. The emotion based combination of music and photos was compared to a Microsoft Media Player visualization and a random photo slideshow. The user test reveals an increased user experience and acceptance of the mood based combination of music and photos. Therefore a cross-modal application of multimedia items should be applied to the TV broadcast generation which implies the best user experience instead of processing audio and visual data individual.

3. ENRICHMENT OF INTERACTIVE TV SERVICES

3.1 Interactive Digital TV Environment and System Overview

The scenarios of the presented paper are realized by developing server components within a digital TV playout environment. The base content and playout server architecture of our system is described in [17]. Core of the architecture is a streaming oriented data exchange between core modules e.g. MPEG-2 transport stream multiplexer, video encoder and object carousel generators. Furthermore, a content, scheduling and administration server controls the complete system and the interaction between the components.

Within the ongoing work for the enrichment technologies special modules for media acquirement, multimedia analysis and collaborative filtering will be integrated, see figure 3. The depicted processing chain shows an abstract workflow for interactive TV generating services. It includes the user message handling which could be mobile phone messages as well as other return messages such as email. The user messages can be filtered for registered users to enable the collaborative filtering module which search for cross-interests to other registered users. Within the next step a media acquiring is processed if needed for the selected service. Afterwards, the content-based analysis module processes audio, image and video data for semantic indexes and similarity. The last module of the enrichment chain realizes the scheduling and preparation of multimedia items also utilizing content-based analysis methods. The workflow can differ by the kind of application.

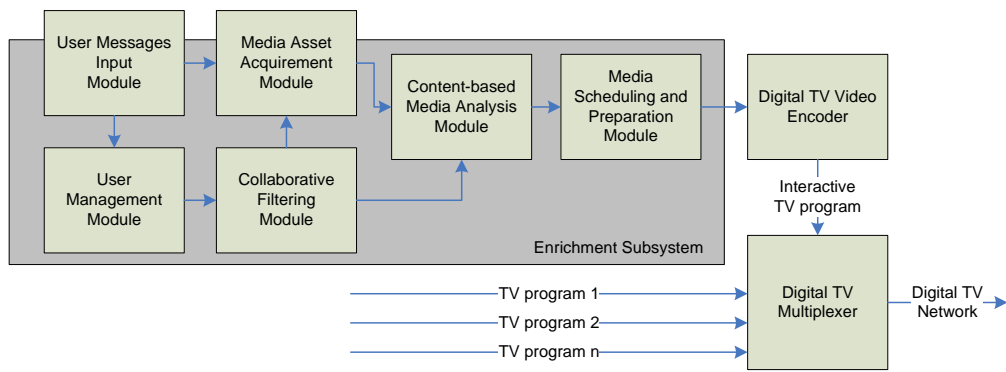


Figure 3: Interactive content and playout server architecture for enrichment of interactive TV services with content-based analysis and collaborative filtering methods.

3.2 Interactive TV Service Enrichment Strategies

Based on the pre-defined system architecture and the depicted content-based and collaborative filtering methods multiple strategies for enrichment of interactive services are possible.

- Text messages of users can be analyzed for keywords to find matching multimedia items within user-generated or automatic annotated multimedia databases to integrate these content in the broadcast.
- Multimedia messages of users can be used as query in content-based retrieval systems to retrieve similar multimedia items for broadcasting.
- Audio content or spoken voice of users can be analyzed for keywords or matching music songs. The keywords and song titles can be used to retrieve multimedia items for broadcasting.
- A user group can remote control intelligent playlist generators for broadcasting video clips or music by relevance feedback methods.
- Collected media assets can be automatically cross-modal scheduled and prepared for the best user experience.

3.3 Example Services

Following the above described enrichment strategies multiple concrete service scenarios are possible. Exemplarily the services *Intelligent Playlist Generator (IPG)* and *Theme Broadcast Generator (TBG)* described in detail.

Intelligent Playlist Generator

The IPG service describes on the example of music an interactive service for a dynamic playlist generation. The principle service concept can also be applied on the images domain as slideshow or in combination as music-slideshow service.

To start the service a user or the service administrator sets up a so called seed song which is used by a music retrieval algorithm searching a music database to retrieve a first playlist of similar songs. This playlist grows and changes in the further progress of the service. The submission of such a song can be handled by the song title via text message or via mobile phone music identification as described in section

2.3. The identified song is delivered to the playlist generator. After the initialization of the service, users can give feedback to currently played songs like voting. A relevance feedback algorithm benefit of the continuous user feedback and adapts the IPG. As worst case of a mass decline the current played song is stopped and an alternative song is chosen. The voting results can be shown on the screen to extend the user experience. Next to the voting approach additional songs can be submitted for a more direct control of the IPG by multiple seed songs and playlist aggregation algorithms. This method can be extended to a multiplayer-game like service: "Find and submit different songs of one special music genre up to the IPG plays your genre". Alternative the collaborative filtering module can be enabled to incorporate already existing music preferences of registered users.

Regarding the music-slideshow version of this service additional techniques for merging audio and visual content e.g. mood classification are applied. Furthermore, audio beat analysis is applied to change images on music beats. Song structure segmentation is applied to repeat the same images within each chorus of a song.

In principle the voting approach is feasible for mass interaction while the submission approach should be used with a limited number of users.

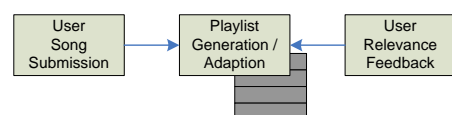


Figure 4: Interactive playlist generation.

Theme Broadcast Generator

The theme broadcast generator service describes an interactive TV service which generates a TV broadcast based on available multimedia content retrieved by keyword search. The service starts by setting up initial keywords for special theme e.g. "New Zealand". The initial keywords can set by a pre-registered user via text message or a service administrator. Incorporating these keywords a multimedia search on different multimedia platforms e.g. Flickr.com is performed to retrieve initial content for the TBG. Next to the multimedia content, knowledge databases like Wikipedia.org are

searched for textual content which can be used with text-to-speech synthesis as audio content. Additional music content can be chosen e.g. by a mood classification of the collected images and a mood-based search in tagged or semantic indexed music databases.

Combining and scheduling collected and generated multimedia data, a theme broadcast can be generated without any production costs except possible license fees. The last step concentrates on the design of transition effects e.g. between images. Therefore, also the above described mood classification as well as pre-defined transitions in mood profiles can be applied e.g. calm and melancholic suit to dissolves.

While broadcasting the initial collected content, users can send text messages with additional theme matching keywords e.g. for the "New Zealand" example: "milford sound, queens town, abel tasman". These interactively provided keywords enrich the original theme broadcast by additional multimedia content which is dynamically integrated. To encourage users to attend to the TBG service nicknames or images of the users can be shown on the broadcast together with the content the user is responsible for.

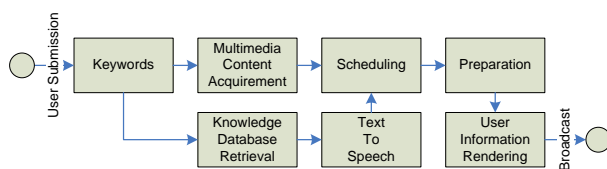


Figure 5: Workflow of theme broadcast generator.

4. CONCLUSIONS AND FUTURE WORK

Within this paper we presented a novel approach for enrichment of mobile messaging services with content-based and collaborative filtering methods. We described the ongoing work of our content and playout system architecture. We discussed content-based analysis algorithms and their possible use in enrichment components. Two service concepts were described to explain the application of the content-based algorithms within an interactive TV broadcast generation process. Summarizing the results a base system for prototypic interactive services was realized.

A crucial aspect of future work is the design and elaboration of additional service concepts on the base of the described enrichment strategies. Finally, field tests with a group of users are planned to evaluate the user experience regarding entertainment, usability and acceptance.

5. ACKNOWLEDGMENTS

This work was partly done in the iKabel project, supported by the Thuringian Ministry of Education and Cultural Affairs.

6. REFERENCES

- [1] Y. Beyer, G.S. Enli, A.J. Maaso, and E. Ytreberg. Small Talk Makes a Big Difference: Recent Developments in Interactive, SMS-Based Television. *Television & New Media*, 8(3):213, 2007.
- [2] Samuel Miller. Taking on the masses with mobile messaging TV. *Computer. Entertain.*, 3(2):6–6, 2005.
- [3] J.L. Herlocker, J.A. Konstan, L.G. Terveen, and J.T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1):5–53, 2004.
- [4] Arnold W. M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.
- [5] A. Bosch, A. Zisserman, and X. Munoz. Scene classification via pLSA. *European Conference on Computer Vision (ECCV)*, 4:517–530, 2006.
- [6] David G. Lowe. Object recognition from local scale-invariant features. *International Conference on Computer Vision*, 2:1150–1157, 1999.
- [7] O. Hellmuth, E. Allamanche, J. Herre, T. Kastner, N. Lefebvre, and R. Wistorf. Music genre estimation from low level audio features. *Proceedings of the 25th International AES Conference*, pages 205–212, 2004.
- [8] Lie Lu, D. Liu, and Hong Jiang Zhang. Automatic mood detection and tracking of music audio signals. *IEEE Transactions on Audio, Speech & Language Processing*, 14(1):5–18, 2006.
- [9] Christian Uhle. Tempo induction by investigation the metrical structure of music using a periodicity signal that relates to the tatum period. *1st Music Information Retrieval Evaluation eXchange (MIREX), London, Great Britain*, 2005.
- [10] J. Aucouturier. *Segmentation of Music Signals and Applications to the Analysis of Musical Structure*. PhD thesis, Master Thesis, King's College, University of London, UK, 2001.
- [11] Christian Dittmar, Christoph Bastuck, and Matthias Gruhne. A Review of Automatic Rhythm Description Systems. *Conference on Music Communication Science (ICOMCS), Sydney, Australia*, 2007.
- [12] Christoph Bastuck and Christian Dittmar. An Integrative Framework for Content-Based Music Similarity Retrieval. *34. Deutsche Jahrestagung fuer Akustik (DAGA), Dresden, Germany*, 2008.
- [13] Kay Wolter, Christoph Bastuck, and Daniel Gärtner. Adaptive User Modeling for Content-Based Music Retrieval. *6th International Workshop on Adaptive Multimedia Retrieval (AMR), Berlin, Germany*, 2008.
- [14] Peter Dunker and Matthias Gruhne. Audio-Visual Fingerprinting and Cross-Modal Aggregation: Components and Applications. *12th International Symposium on Consumer Electronics (ISCE)*, 2008.
- [15] Peter Dunker, Stefanie Nowak, André Begau, and Cornelia Lanz. Content-based mood classification for photos and music: A generic multi-modal classification framework and evaluation approach. *International Conference on Multimedia Information Retrieval (ACM MIR), Vancouver, Canada*, 2008.
- [16] Chin-Han Chen, Ming-Fang Weng, Shyh-Kang Jeng, and Yung-Yu Chuan. Emotion based music visualization using photos. In *14th International Multimedia Modeling Conference*, 2008.
- [17] Peter Dunker, Uwe Kühhirt, and Andreas Haupt. A System for Enhanced Services in CATV Networks. *9th Workshop Digital Broadcasting, Erlangen, Germany*, 2008.