Mobile Video as an Instrument in a Social TV Service

Minna Suovirta
VTT Technical Research Centre of Finland
P.O. Box 1000
02044 VTT, Finland
+358 20 722 5916
minna.suovirta@vtt.fi

Markus Ylikerälä
VTT Technical Research Centre of Finland
P.O. Box 1000
02044 VTT, Finland
+358 20 722 6082
markus.ylikerala@vtt.fi

Chengyuan Peng
VTT Technical Research Centre of Finland
P.O. Box 1000
02044 VTT, Finland
+358 20 722 6839
chengyuan.peng@vtt.fi

ABSTRACT
Video derives as a larger part of social media its value from active participation of services with opportunities for social interaction and engagement. Into social video services can be integrate other communication services like voice, chat, context awareness, and ratings to support a shared TV experience within a community. We make a research and pilots to find out how mobile video can effectively be utilized in service provision and increase mobile TV experience.

In this paper we describe our vision and a platform of a mobile social video service. We show with three pilots the social video concept which can provide a great opportunity for content providers to keep directly in touch with users. The system also uses metadata based on semantic web for social discovery of video content generated by mobile users and for adapt content to what users want.

Categories and Subject Descriptors
H.1.1 **[Models and Principles]**: Systems and Information Theory – *Value of information*. H.3.3 **[Information Storage and Retrieval]**: Information Search and Retrieval. H3.4 **[Information Storage and Retrieval]**: Systems and Software – *Information networks*. H.3.5 **[Information Storage and Retrieval]**: Online information services. I.2.7 **[Artificial Intelligence]**: Natural Language Processing – *Speech recognition and synthesis.*

General Terms
Performance, Reliability, Experimentation

Keywords
Mobile Social Video, user generated content, metadata

1. INTRODUCTION
The television industry is facing one of its biggest changes in the form of personal computers, broadband Internet, and mobile. There is an on-going rapid expansion of video contents on the Internet and progress towards the "3-screen TV" model in reference to the growth of video on both PC and on mobile devices. There is plenty of room to expand the social interaction and participation opportunities for video content. This calls for innovation in developing novel social television services. There will also be plenty of possibilities in new mobile services, including those users who will become co-creators. The mobile social video will be integrated and interact in complex ways with other forms of video use.

Professional content providers and individual consumers will work in new interactive co-operation in the future of social video. The new systems like social video will help individuals and communities to publish fast and good quality video content with other individuals but also in co-operation with professionals. The roles of the players are changing so that consumers will be important content providers and professional companies will have new responsibilities and also new business opportunities.

Co-creative video production activities will be based on the collective power of many creators and call for new solutions on media metadata, real-time annotation and cross-linking of media sources, automatic media composition, innovative media formats, and new solutions for handling of intellectual property. This means call for novel tools and services that provide easy tools for capture and sharing, context-awareness in video creation, support for co-creativity among users, and direct integration into media production and distribution infrastructure.

Our social video concept is general and independent of industrial domain and it boosts services for third parties, whose business may have nothing to do with video itself. We are considering both content and technology. Content can be everything from mass media by professional creators to user generated media. Similarly technology aspects are not tied to a certain network or distribution mechanism. However, the focus is on video produced by individuals and communities in co-operation with professionals.

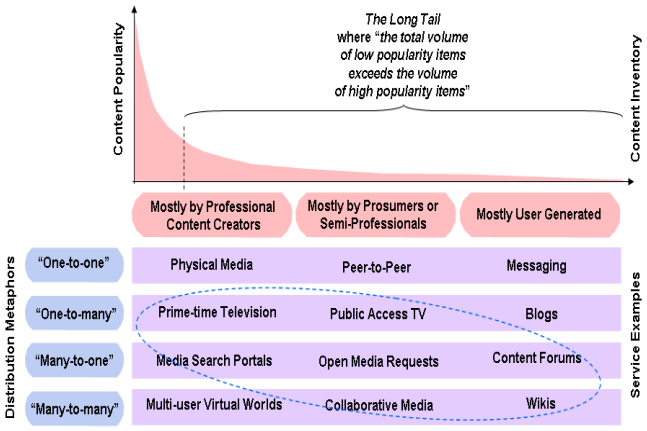

Figure 1. Scope of the project

Scope of the project [1] is shown in Figure 1. The base to social video service concept development is to find out (user and other) requirements for next generation collaborative social video production. We research and develop both technology and user related aspects.

In a social video service should be something that motivates users to use and contribute videos. Project partners'[1] knowledge and technologies are used to make an integrated social video concept in the multi channel environment (cableTV, IPTV, mobileTV, localTV) with hopes to offer good user experience.

In order to individual videos to be usable we need some ways to select, validate and package those. Our main point in this field is to identify false information from user generated content and select, filter and collect information from different sources to useful packages for different social video usages.

In the service concept development common standards for meta-data and interfaces play important role in successful results. The current major problem of the semantic search is the lack of available semantics for the resources. Users should have different easy ways to inject keywords to be used as metadata.

2. ARCHITECTURE AND PLATFORM

In this section we present our system and platform architecture.

The SocialVideo Distilled is a location based service (LBS) that passes videos into possession of location for collectively improving available videos with user generated metadata. Figure 2 shows the architecture of the project that basically consists of video and metadata management from heterogeneous sources, and tracking of the user based on the user's mobile device. The user interface on the mobile device can be used for user based metadata generation and broader delivery of the improved distilled videos.

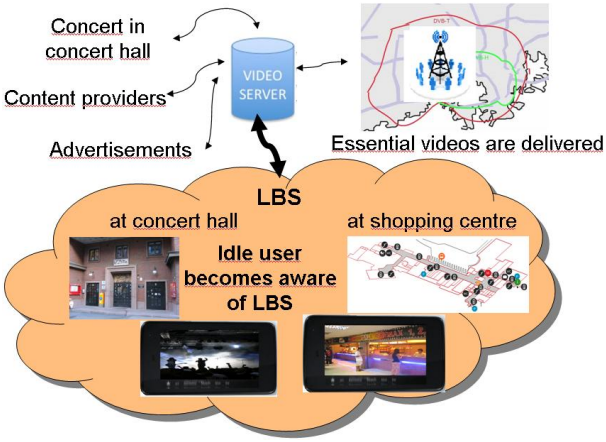

Figure 2. Social Video Distilled as LBS

The uSpace [2] provides a platform for the actual service i.e. SocialVideo Distilled. Figure 3 describes the mobile platform building blocks based on uSpace. The uSpace evolved within LUCRE [3] service development kits is a collaborative and ubiquitous service engine. It extends current concept of widget containers, e.g. iGoogle and Netvibes. End-users can create and manage dynamic web services combining a set of widgets from a library to a shared space. It also aims for spatially and temporally aware services which treat location and time as metadata.

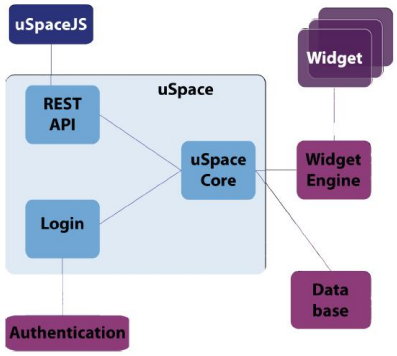

Figure 3. Architecture of the uSpace and the relating components [2].

Thus, a logical video server contains arbitrary multivendor videos such as from concerts as well as from advertisers or combined with advertisements. A user on the location (e.g. the concert hall) becomes aware of the service based on the tracking of user's mobile device by e.g. GPS, WLAN, Bluetooth, etc. The user can further access the service with the mobile device. Next, the service provides the previously recorded videos at the concert hall for the user to comment, grade etc and the user's contribution is finally included to the service.

3. METADATA USAGE

An essential aspect of these approaches is the inclusion of metadata. Metadata is essential for helping recommendations, search and discovery. The service can keep control of the tags and metadata and link users to content via a specific event. Semantic search of content could aggregate video from several sources, personalize them to the user interests and location, and to augment the probabilities for content to be found faster and with more relevance to users.

We designed an end-to-end metadata architecture for semantic and collaborative search of content shown in Figure 4. The central part is the deployment of Drupal content management framework [4]. The content metadata (knowledge) store is a structured collection of records of metadata. We use the Resource Description Framework-in-attributes (RDFa) as metadata format [5]. This architecture is responsible for serving xhtml web pages with embedded RDFa metadata queried by uSpace platform. This usage architecture can build a distributed and collaborative editorial model in broadcaster site.

The modules in Figure 4 (in next page) including the third party modules are functional plug-ins that can interact with the dynamic content databases created by end users. The process of configuring RDFa support is simple and straightforward.

[1] Arcada University of Applied Science, Digita Oy, Elisa Oyj, Forum Virium Helsinki, Helsinki Institute for Information Technology (HIIT), Lingsoft Oy, Nokia Oyj, Sofia Digital Oy, Sanoma Entertainment Oy, Sanoma Television Oy, Teosto ry, University of Tampere, Tampere University of Technology, TVkaista Oy, Vincit Oy, VTT Technical Research Center of Finland

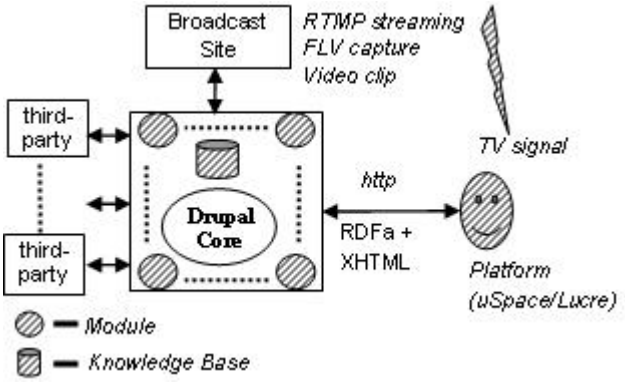

Figure 4. Metadata Flow Diagram.

In the project we use the concept of metadata bundle, that is, in addition to the base video metadata element set (cf. table 2), other optional metadata sets can be combined for different scenarios (cf. table 3, contextual metadata elements).

Table 2. Base Metadata Element

Attribute	Vocabulary	Description
title	DCMES	title of the video file
creator	DCMES	creator of the video file
date	DCMES	date that the video was uploaded
identifier	DCMES	URL of the video file
subject	DCMES	a list of keywords describing the content of the video
format	DCMES	format of the video file
rights	DCMES	a URL to a text-based description of the video's licensing terms
description	DCMES	provides a text-based description of the video
relation	DCMES	relationship of this resource to another resource
source	DCMES	objects, either electronic or printed, from which this resource was derived
type	DCMES	the nature or genre of the content of the resource
extent	DCMI	the size or duration of the resource

The Simple DCMES (Dublin Core Metadata Element Set) consists of 15 metadata elements, from which we found main part suitable for us. The additional terms are identified, generally in working groups of the Dublin Core Metadata Initiative (DCMI).

Table 3. Contextual Metadata Elements

Attribute	Vocabulary	Description
position	GeoTag	latitude;longitude position coordinates for location information about the current location of the device
placename	GeoTag	the name of the place
region	GeoTag	the name of the region, country
weather, temperature, humidity	SensorML	Current environmental conditions including weather, temperature, and air conditions
atDateTime/created	Timeline Ontology /DCMI	the current system time of the content consumption

4. PILOTS

We have three pilots in different areas. There are many interesting questions to which we hope to get useful answers via our pilots. We among others think issues relating metadata, user-generated content, editing and users of the video content.

Users of mobile phones are beforehand recruited and advised to use a test mobile phone with appropriate applications. User tests and interviews are performed both during the development process and in the pilots in order to contribute social service development.

4.1 Live concert

We have couple of live concert user tests to get needed mass of people and variation of concert situations to get reliable user feedback results. We use both an automatic and a manual remix, so that we can compare implications and the social activity the different editions create. The concert pilots are arranged on April 2010.

In a band gig several fans record video clips by mobile phone. After the gig fans upload their clips into a social video service. A remix video, made from automatically selected video clips, is played on music club's widescreen right after the gig. The video clips are also put available to fans for manual remixing. Both editions are put available for the fans to view and discuss in the band's web page.

In a music hall are series of band competition concerts. Concerts will be recorded both by mobile phone cameras and professional equipments. Mobile phone shooters change in the concerts so that we get feedback from various test users. Mobile phone users upload their video clips into a server and an automatic video clip remix is also done in that case so that broadcast editors can use it in remixing professional multi-camera material and user-generated material into a live video entity. In that way we get results how professional broadcast content can be enhanced with user generated content during a live event. After the edition the concert materials are delivered in multiple channels to test multi channel environment.

To the social video service can be included speech recognition as a way to users to give and system to find out metadata. Speech recognition is much easier from video clips taken in a less noisy event, so we test speech recognition as a source of metadata partly separately to above mentioned pilot cases.

In both cases it can be seen that all interested parties (the band/, band's fans and the music club/hall, etc.) have their own benefits and interest to promote the social video production.

4.2 Added value broadcast

Traditionally friends, families and other communities get together and gather around television to watch e.g. live sports event or a TV series. They discuss and communicate with each other at least during the event or an episode.

The social aspect of gathering physically together can be transferred to web environment by allowing commenting and sharing video messages and synchronized feed with each others during a live event. Other synchronized video material is streamed to the community (an open/enclosed auditorium in Facebook), no matter where the participants physically are as long as they have the internet connection. The communal and social video viewing

is transferred to a web site by means of a computer or mobile phone.

These communities extend far beyond the home to span entire neighborhoods, cities, countries, and hemispheres. And like the traditional living room, these communities are increasingly organized around video, connecting families, friends, and some strangers alike in a shared video space defined by interactions, common interest, or location.

Two sessions of the pilot are arranged on April. Test users gather together to view same sport broadcast and use system to communicate each other. The pilot tells us, how well social commenting possibilities and user generated contents bring added value to the professional broadcast during a live event.

4.3 Case of Stories

Group of students are given a task to record five minutes of their life every day for a week concentrating on a given theme. The students can include in beginning of the video clips some optional predefined keywords e.g. 'place school' (in Finnish) to be detected by speech recognition. Pilot tests are arranged in beginning of May. The resulting videos are tagged, archived, and edited into a story which can be used among other content in delivery of multiple channels.

The goal of the pilot is to show that user generated social content can make a big story. The case is also part of StadiTV [6] project.

5. CONCLUSIONS

There are many drivers to use a video as a social media instrument. Users are used to take videos by mobile phones and video cameras, and send clips to relatives and friends. Why don't send conveniently and fast a descriptive video than write same message e.g. in a letter?

There is huge potential to user generated content creation in itself and as enhancement of a professional media. The professional media actors has pressures to utilize user generated content and expectations social video to be a new way to engage and motivate users to be their customers.

A video presentation may tell much, but a file name automatically generated by a digital camera does not. To search, find, group, etc. video materials you need to have metadata with video files. This calls for good video services to save materials.

A video service could find automatically part of the basic metadata e.g. time, but all significant metadata can not be got automatically and a machine may misinterpret. So, the users should get motivated to add metadata. One possibility is to use automatic speech recognition in giving and finding out metadata.

We test speech recognition in our pilots in giving metadata and taking part in a social session within a video. We expect good results, if background noise is not too loud. After the pilots we can say more about acceptance of the speech recognition feature among the test users.

Project has got good results in automatic detection of e.g. time and location metadata from video data, which is created within the project or linked e.g. from YuoTube. We are also able to substantially identify false information and cluster videos to useful packages for different social video usages.

In the pilot cases are various motivating issues for users to use and contribute social videos in the service. After the pilots we can say more about motivation issues.

The user generated content as well as professional one should be edited. Everybody does not want, has time or can do that. So, it would be nice to put own clip to a service and in addition to see others' clips e.g. from the same concert and get even a remix of whole event. After the pilots we can say more about successfulness of that scenario.

We deliver video content within copyrights in multiple channels. After the pilots we can say more about successfulness of multi channel delivery and utilization of same materials. Copyrights affects to it, what is possible and allowed to do with individual video materials. Material authors (composers, bands, etc.) put it own challenges to utilization of materials. Copyrights to a one-time-use in e.g. a concert hall's wide screen can be quite easily arranged, but how about in case of e.g. on-demand multi-channel delivery? User generated content and their materials delivery via many different channels challenges copyright organizations to license repertoire development and further international collaboration between actors in different countries.

Current pilots are our ongoing work. We will present detailed results after. Our main contribution is the design of a video-oriented mobile social TV service system in which various social features can be technically integrated with the actual mobile TV viewing experience.

Current mobile TV services could certainly benefit from this augmentation - from the traditional uni-directional viewing screen to a more connected and shared space with family, friends, and communities. Semantic technologies used also make it possible to give richer descriptions to mobile video, facilitating the process of locating, combining diverse media from various sources and personalizing content recommendation.

6. Acknowledgement

The authors would like to thank project participants for their inspiriting visions and fruitful work to develop a concept and platform for a mobile social video service. Special thanks to Arcada, Forum Virium, HIIT and Sanoma for opening up good pilot case topics.

7. REFERENCES

[1] Social Video, http://www.flexibleservices.fi/en/node/32.

[2] uSpace, http://www.flexibleservices.fi/files/file/pdf/LUCRE_D25_ReleasesOfLUCREToolsAndSystem.pdf

[3] LUCRE (Local User Created Services), http://www.flexibleservices.fi/en/node/28

[4] Drupal, http://drupal.org

[5] RDFa, http://www.w3.org/TR/rdfa-syntax/

[6] StadiTV, http://www.forumvirium.fi/stadi.tv-2.html