What do you wish to see? A Summarization System for Movies based on User Preferences

Rajkumar Kannan[1*], Gheorghita Ghinea[2], Sridhar Swaminathan[3]

[1] King Faisal University, Al Ahsa 31982, Saudi Arabia

rkaruppan@kfu.edu.sa, Tel: +966-13-5899273

[2] Brunel University, Middlesex UB8 3PH, United Kingdom

george.ghinea@brunel.ac.uk

[3] Bishop Heber College (Autonomous), Tiruchirappalli 620017, India

sridarah@gmail.com

* Corresponding author

Abstract. Video summarization aims at producing a compact version of a full-length video while preserving the significant content of the original video. Movie summarization condenses a full-length movie into a summary that still retains the most significant and interesting content of the original movie. In the past, several movie summarization systems have been proposed to generate a movie summary based on low-level video features such as *color*, *motion*, *texture*, etc. However, a generic summary, which is common to everyone and is produced based only on low-level video features will not satisfy every user. As users' preferences for the summary differ vastly for the same movie, there is a need for a personalized movie summarization system nowadays. To address this demand, this paper proposes a novel system to generate semantically meaningful video summaries for the same movie, which are tailored to the preferences and interests of a user. For a given movie, shots and scenes are automatically detected and their high-level features are semi-automatically annotated. Preferences over high-level movie features are explicitly collected from the user using a query interface. The user preferences are generated by means of a stored-query. Movie summaries are generated at shot level and scene level, where shots or scenes are selected for summary skim based on the similarity measured between shots and scenes, and the user's preferences. The proposed movie summarization system is evaluated subjectively using a sample of 20 subjects with eight movies in the English language. The quality of the generated summaries is assessed by *informativeness*, *enjoyability*, *relevance*, and *acceptance* metrics and *Quality of Perception* measures. Further, the usability of the proposed summarization system is subjectively evaluated by conducting a questionnaire survey. The experimental results on the performance of the proposed movie summarization approach show the potential of the proposed system.

Keywords: Video summarization, movie summarization, video semantics, personalization, user preferences.

1. Introduction

The past decade has witnessed an explosive growth of digital videos, both over the Internet and on home computers. However, browsing through lengthy and voluminous video collection becomes tedious to the user, if the videos have little or no relevant content. Moreover, searching for interesting segments within the videos is time consuming [16].Video summarization systems generate a compact version of a lengthy video and help the users to watch its

significant segments [6, 29]. According to the types of content used for video analysis, existing video summarization methods can be classified into *cognitive-level* approaches and *affective-level* approaches [29]. The cognitive-level video summarization approaches [7-11] extract low-level features such as *color, motion, texture, and audio*, *visual* and *textual* saliencies to identify important segments of a video. On the other hand, the affective-level summarization approaches [12-14] generate video summaries by modeling the affective video content by exploiting users' feedbacks/responses while watching a video. In general, most of these approaches extract significant segments of a video based on low-level features that are considered to get the users' attention. However users always desire video summaries that are generated based on the high-level features such as events, semantic concepts etc., rather than low-level features alone [18]. It is rather unsurprising, therefore that the well-known semantic gap problem is thus shown to also exist in video summarization.

Most of the existing video summarization systems are generic in nature. That is, for a video, these systems create a video summary that is common for all the users. Generic video summarization will not be sufficient when the users' needs and interests differ and change over a time [18]. Thus, the users are seldom satisfied by a *generic video summary* (also called *non-tailored video summary*) produced by a video summarization system [18]. This is because the produced video summary may not contain video content of the particular event, semantic concept or genre liked by the user. A generic video summarization system that produces video summaries based only on low-level video features cannot process a semantic level query from a user, such as '*summarize all the action events of a movie*'. Personalized or tailored video summarization is a useful technique for producing tailored video summaries to the users based on their needs and interests [18, 19]. Also, recent information retrieval and filtering systems have started tailoring or personalizing the results by adjusting to individual user's needs and interests. Accordingly, we argue that the criterion used to summarize a video should be the user's preferences and interests over the high-level features of a video.

Film is an art form that offers a practical, environmental, pictorial, dramatic, narrative and musical medium to convey a story [1]. Trailers, which are short summaries of movies, have been used for decades to promote movies. Creating movie summaries manually by domain experts is a tedious and time consuming task as the users' viewing time constraints and preferences change over a time. Movie summarization, being a special class of video summarization, is particularly challenging since a large variety of movie scenarios and film styles complicate the summarization problem [29]. Movie summaries help the user to decide whether to watch an entire movie or not. Most of the movie summarization methodologies generate summaries based on the visual attention and motion activities in a movie [8, 10]. These summaries will be semantically meaningful only if the summarization methodology considers high-level movie features [2]. Since characters are considered as the important high-level features of a movie, recent approaches for movie summarization [28-30] identify roles in a movie by exploiting social network analysis and role recognition. Apart from characters, users might also be interested in important events, and semantic concepts in a movie for summarization.

Movie videos contain rich high-level features such as *characters*, *events*, *semantic concepts* and *spoken content*. The flourishing movie industries produce more than 4500 movies every year [30]. Manual annotation of the aforementioned movie features for a large number of movie videos is tedious, time consuming and laborious. This necessitates movie summarization approaches to exploit efficient indexing and automatic annotation techniques for preprocessing. Video Content Analysis (VCA) greatly helps the users for an effective media management including indexing, retrieval, and summarization. Bridging the "semantic gap" has always been one of the notorious and biggest challenges in video content management,

i.e. allowing the users to browse, retrieve, and summarize video content at semantic level [3]. Even though movie content analysis does not always need a real-time processing as required for surveillance, and sports video analysis, the rapid growth of movie videos necessitates an efficient movie content analysis. Thanks to the recent advancements in the fields of Computer Vision, and Pattern Recognition in high-level feature detection, and recognition which has let the Multimedia Information Retrieval (MIR) applications to exploit faster and accurate automatic annotation techniques.

As shown in Fig. 1, the granularity of movies goes from video frames, shots to scenes and substories. A video should be broken down into a set of segments which are either shots or scenes for efficient indexing. *Shot Boundary Detection* (SBD) [4] and *Scene Change Detection* (SCD) [5] techniques temporally decompose a video into more manageable units. Also the manual efforts for annotation of shots and scenes can be reduced by automatic video content analysis techniques. Moreover, *face recognition* [39] and *large scale semantic concept detection* [35] techniques can be utilized for annotating visual features of a movie automatically. The spoken content of a movie can also be automatically annotated using *Automatic Speech Recognition* (ASR) techniques, or *closed captions* that are available in the web as subtitles. In personalized movie summarization, users may be interested in all possible high-level movie features. User profiles, as employed in recommender systems, can be utilized for such a personalization task.

Unlike most of the generic video summarization approaches which summarizes a video based on the objective criterion such as saliencies, user feedbacks/responses, character analysis, information coverage, and diversity, the proposed summarization methodology generates semantically meaningful personalized movie summaries by exploiting user's subjective preferences over the high-level visual and textual features of a movie. The novelty of the proposed approach as opposed to other movie summarization and personalized video summarization approaches is that, the proposed approach supports multiple real valued preferences over a wide varieties of high-level movie features for both shot level and scene level movie summarization. Also, the proposed approach generates personalized summaries in a unified manner using effective similarity measures, prioritized fusion of different semantic level similarities and a constrained selection scheme.

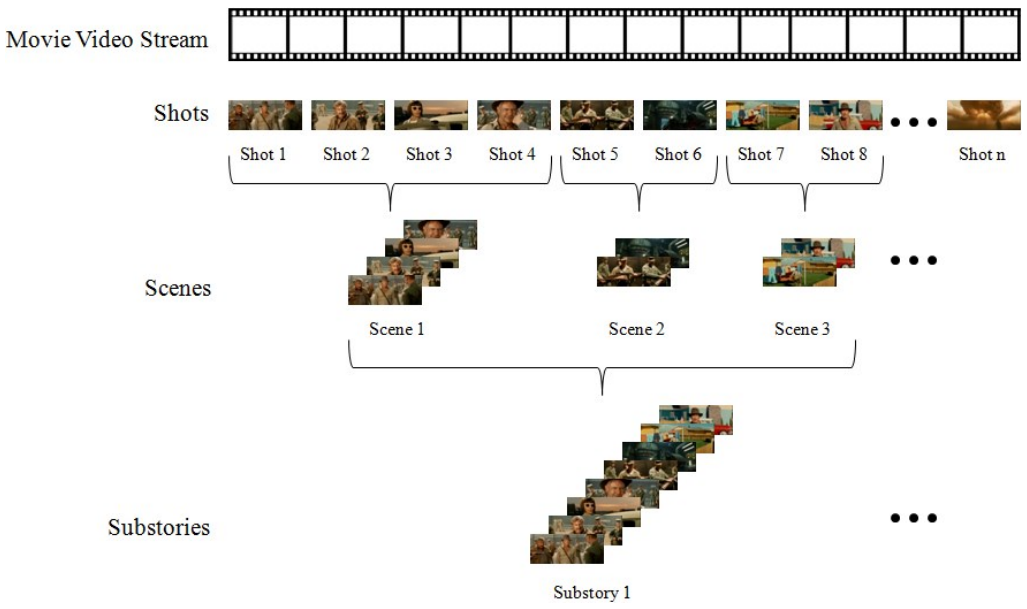

Fig. 1 Granularity of a movie from video stream to substories

Our hypothesis is that, a single generic movie summary does not satisfy every user, and summaries for the same movie should be generated based on individual user's preferences. Hence, this paper proposes a novel system for personalized movie summarization that produces tailored movie summaries by adapting to the user's preferences. The following are the contributions of this paper.

- A novel personalized movie summarization methodology that generates semantically meaningful personalized movie summaries both at shot level and scene level by exploiting individual user's preferences in a movie which are obtained from a query interface.
- A user study based on Quality of Perception measures demonstrating the advantages of personalized movie summarization over generic movie summarization, and a subjective study on the system usability that shows the subjects' diverse behavior regarding the usability of the proposed movie summarization system. The subjective user study also reveals the challenges in developing a new and more sophisticated personalized paradigm for movie summarization.

The rest of the paper is organized as follows. Section 2 reviews the related work on video summarization and in particular movie summarization. Section 3 describes the overview of the proposed movie summarization system. The proposed methodology for personalized movie summarization is presented in Section 4. Experiments and results focusing on subjective evaluation of the proposed movie summarization technique and of the system usability are reported in section 5. Section 6 concludes the paper and opportunities for future work are identified.

2. Related Work

A wide number of approaches have been proposed over the years for video summarization. Generally, video summarization systems are classified into two types based on the summary visualization methods - *keyframes* and *video skims*. Keyframe visualization does not convey much information to the users because it lacks the temporal property of a video. Video skim comprises a collection of meaningful video shots, which is considered as a simple yet powerful visualization method for video summarization. A comprehensive survey on the recent video summarization techniques can be found in [6].

2.1. User Attention based Video Summarization

Various user attention models have been proposed for summarization to make use of the users' perceptual response to low-level audio, visual and textual features [7-11] (cognitive-level approaches) and the users' response while watching a video [12-14] (affective-level approaches). Audio, visual and linguistic attention models were used to generate the attention curve of a video for both static and dynamic video summarization [7]. The authors utilized both low-level attention models such as motion, static, camera and audio attention models and mid-level attention models such as face, speech and music attention models. A multimodal saliency curve is constructed for movie summarization by integrating audio, visual and textual saliency curves of a movie video stream [8]. It uses a spatio-temporal saliency model, an AM-FM speech model and Part of Speech (POS) tagging for computing visual, aural and textual saliencies respectively. Video features that easily attract users' attention and influence

human perception, such as motion, contrast, special scenes and statistical rhythm, are extracted and modeled for summarization [9]. Attention scores are computed and attached to the scene, clusters, shots and subshots in a temporal graph for video summarization [10]. The authors used a motion attention model for computing the visual attention scores. Importance of a video segment is determined using term, document frequencies and bigrams in the speech transcripts of a video [11]. Summary is then generated by selecting most informative segments of the video.

Most of the attention (or saliency) based cognitive-level video summarization methods utilize users' task-independent response or attention to audio, visual and textual modalities of a video. The advantage of using attention based approaches is that, they are computationally efficient enough for real-time video summarization. A major limitation with cognitive-level approaches is that, they often fail to work well on videos with semantically rich content (like movie and sports videos). However, the summaries generated based on saliencies might not contain semantically interesting or significant content of a video, as they do not consider high-level video features. Thus, it is necessary to consider the semantics underlying a video and the users' requirements for goal-oriented, task-specific video summarization.

Variations in user's eye movement, blink, and head motion are considered for identifying interesting segments of a video [12]. The authors in [13] presented an affective video summarization approach based on the facial expressions of viewers while watching the video. Facial expressions were analyzed to infer affective scenes from videos. Affective segments of videos and Regions of interest (ROIs) are discovered by analyzing the viewers' eye-gaze [14]. An advantage of using affective-level video summarization approaches is that, they implicitly utilize users' interests and responses while watching a video. However, the effectiveness of these methods often depends upon the ability to capture users' responses and mapping of such responses to the corresponding video segments. Also, cognitive-level approaches always need controlled summarization setups. Same as the attention based video summarization methods; these methods do not consider high-level video features and users' requirements and thus suffer from less generalizability.

In order to overcome the aforementioned limitations in the context of movie summarization, the proposed summarization system employs a unified approach for personalized semantic movie summarization.

2.2. Personalized Video Summarization

High-level video features are used as preferences to the users for personalized video summarization [15-20]. IBM research has proposed a personalized video summarization system for pervasive mobile devices such as PDA [15]. User, device and transmission profiles were used for adaptive personalized video summarization and transmission. However, their system allows only a single visual semantic concept as binary preference at a time. The importance of a video segment is measured using user's constraints and preferences over audio-visual semantic concepts [16]. Users' Degree of Interest on event, person, and object were used for personalized summarization of life-log videos in a multi-camera office environment [17]. This approach totally relies on manual annotation of events such as *working*, *eating*, *printing*, *meeting*, etc. Semantic concepts such as *humans*, *explosion*, *indoor*, *outdoor*, *close-up*, *zoom-in*, *moving objects*, etc. were automatically detected from videos for personalized summarization [18]. The authors used a constrained optimization problem for selecting shots that are relevant to the user's preferences. Same as their method the proposed summarization also uses constrained optimization for selecting shots and scenes that are relevant to individual us-

er's preferences. Our previous work in [19] proposed a personalized video summarization methodology for summarizing a video based on users' preferences on a set of semantic concepts.

Most of these approaches explicitly obtain users' preferences for shot level personalized video summarization. A limitation with most of these approaches is that they support only binary valued preferences. Binary valued preferences might not be adequate when the user wants to relatively prioritize the different preferences. In contrast, the proposed system supports multiple real valued preferences at a time. These previous approaches for personalized video summarization provide only a limited number of high-level features as preferences to the users. Also, computation of similarity between high-level features and user's preferences, and the prioritized fusion of similarity scores of different video semantics were marginally discussed in these work. The proposed movie summarization system provides a wide variety of high-level movie features such as characters, events, semantic concepts and keywords as preferences to the users. It uses efficient similarity measures and a prioritized linear fusion of different similarity scores for both shot level and scene level personalized movie summarization.

2.3. Movie Summarization

Various attempts were made over the past decade to automate the summarization of full length movies [21-30]. The *VAbstract* [21] system extracts scenes with dialogue, high contrast and high motion to construct trailers for feature films. It excludes the last parts of the original video to keep the suspense of a movie in the trailer. Hollywood-like movie trailers are automatically created by finding trailer patterns [22], where trailers are enhanced using music, sound effects and 3D animations. Trailer patterns are discovered by identifying the positions of trailer segments in the original movie. Sub-stories from movies are detected using short-term and long-term audiovisual tempo analysis [23]. Movie skim is created from the detected sub-stories based on the users' requirements on summary length. *IM(S)*^{2} [24] system summarizes a movie based on user preferred shots of the movie. Personalized summaries are created by comparing audio-visual features of the preferred shots and rest of the shots in a movie. However, the comparison of user preferred content and actual movie content is done only at feature level, not at semantic level. Comic visualization provided by the system *DigestManga* [25], allows the users to interactively integrate movies with comics using an interface.

Two important summarization criteria - *coverage* and *diversity* are considered for summarizing unconstrained videos, where summarization is treated as a combinatorial optimization problem [26]. The objective is to include most informative and diverse elements of an original video into the summary. Summarizing multi-view videos like surveillance videos was formulated as graph labeling task [27]. The authors considered different summarization objectives, such as minimum summary length and maximum information coverage using graph optimization. Our proposed system lets the user to determine both semantic information to be covered by the summary and diversity among summary content, by allowing user to specify their preferences from a range of semantic elements for a particular movie.

A multi-video summarization approach is proposed [51] to generate condensed, descriptive, and aesthetically pleasing video summary for multiple sightseeing videos. The authors have applied the multi-task feature selection approach to discover the semantically important features from videos wherein video frames that can efficiently reconstruct the salient objects in the videos are chosen as keyframes. Also, the authors in [51] used a probabilistic model for

fitting the keyframes into aesthetically pleasing, and coherent video summaries. In our proposed work, summary construction is carried out using a constrained optimization approach where shots/scenes that are semantically relevant to the user's interest are selected and temporally ordered for constructing a movie summary. Unlike multi-video summarization which identifies visually unique and significant video segments, our movie summarization approach discovers semantically relevant and interesting segments for generating a video summary.

Recent movie summarization techniques [28-30] have started exploiting the characters analysis to construct movie summaries. The method proposed in [28] constructs a role network and identifies all leading roles and role communities for summarizing a movie. Relationships between role-communities in a movie are identified for scene based movie summarization [29]. Also a social network is constructed to characterize the interactions between role-communities. The authors provide three types of summaries as preferences to the users - summary covering more scenes consisting of major roles, summary with motion content, the more the better and a summary focusing on movie endings. Character-based movie summarization approach is proposed in [30] where scripts of movies are used in movie analysis for scene and sub-story detection. Importance of a video segment is computed using scores of character involvement, frequent leading character occurrence and conflict between leading characters.

Most of the approaches for video summarization thus aim at utilizing objective summarization criteria such as audio, visual, and textual saliencies, user's feedbacks/responses, characters analysis, information coverage, and diversity. These generic summarization methods might fail to satisfy the users' diverse subjective requirements for summarization. However, a video summarization methodology should also consider increasing the user satisfaction levels with the summary by adapting to their interest. So the criterion for movie summarization should be individual user's subjective preferences.

To the best of our knowledge there is no personalized movie summarization system that integrates visual and linguistic modalities of a movie at the semantic level. The objective of the proposed movie summarization methodology (i.e. what to be summarized) is determined by the users' subjective preferences. The proposed system generates both shot level and scene level personalized movie summaries which are both informative and satisfactory to the users.

3. System Overview

The proposed movie summarization system consists of three modules: *pre-processing*, *user interface* and *video summarization*. The proposed system can generate tailored summaries for movies of any genre. The architecture of the proposed movie summarization system is shown in Fig. 2.

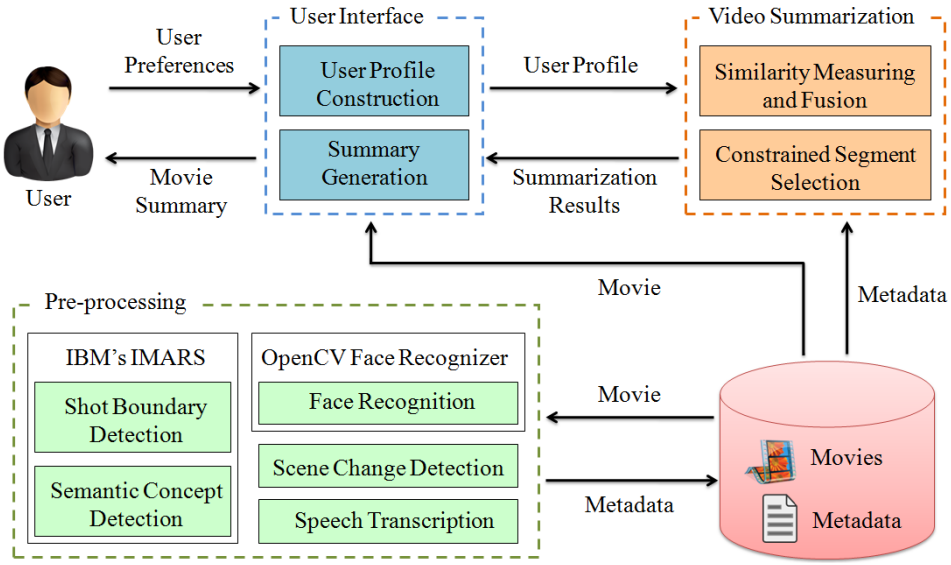

Fig. 2 Architecture of the proposed movie summarization system. The pre-processing module segments a movie video and annotates its high-level visual and textual features. The user interface module constructs the user profile using the user preferences, where the similarity between user profile and metadata of a movie is measured and fused in the video summarization module. In the user interface module, summary is generated from the selected movie segments and presented to the user

3.1. Pre-processing

The proposed system utilizes existing tools and methodologies for pre-processing. As depicted in fig. 2, the pre-processing module segments and semi-automatically annotates high-level features of a movie, and sends these annotations as metadata to the database, which is further used for the summarization. Firstly, a movie is automatically segmented into a set of shots and scenes, where high-level visual and textual features are semi-automatically annotated for each shot. The movie actors, events and semantic concepts are considered to be high-level visual features of a movie. The high-level visual features especially the semantic concepts are manually chosen by considering their frequency of occurrences within the domain of movie videos and their ease of detection using a machine learning classifier. The high-level feature annotation is considered to be semi-automatic, since it comprises a combination of manual and automatic annotation strategies. Visual features such as semantic concepts and actors are annotated automatically, where some movie events are annotated manually. Finally, transcripts of each shot are manually annotated using subtitles of a movie.

IBM's Multimedia Analysis and Retrieval System (IMARS) [31] is used for shot boundary detection and semantic concept detection. Given movie video is segmented into a set of shots based on the low-level visual features. Since a key-frame represents a shot, a single key-frame is extracted from the middle of each shot. Twenty one semantic concepts such as *beach*, *blue sky*, *building*, *flowers*, *greenery*, *people*, *indoors*, *infant*, *landmark*, *mountains*, *nature*, *outdoors*, *party*, *pet*, *skyline*, *sport*, *sunset*, *urbanism*, *vehicles*, *water and wedding*, are detected from each key-frame. The concept lexicon used in this paper is built with twenty one semantic concepts which were manually chosen by considering the *availability* and *detectability* of semantic concepts. As stated in [34], two main properties are required for any desired concept lexicon.

i. The concepts in the lexicon should have high occurrence frequency (i.e. highly available) within the descriptions of real-world images, which makes them commonly used concepts. In our case, the concepts should frequently occur within the domain of movie videos.

ii. The chosen concepts are expected to be visually and semantically consistent (i.e. easily detectable), that is, the images of these concepts have smaller semantic gaps, which make them moderately easy to be modeled for retrieval and annotation. The sematic gap in semantic concepts denotes the difficulty in modeling or representing a concept using the visual features. For example, it is well acknowledged that modeling "Europe" is more challenging than modeling "sunset" due to the lack of an effective visual feature that can represent the concept of "Europe" [34].

The lexicon of 21 semantic concepts used in this work, is a subset of the concepts in Large-Scale Concept Ontology for Multimedia (LSCOM) [32] which contains 449 semantic concepts that are manually annotated on broadcast news videos from the TREC video retrieval evaluation (TRECVID) benchmark. Since, LSCOM lite (subset and lighter version of LSCOM) [33] comprises very common and easily detectable 39 semantic concepts, we have chosen 10 semantic concepts from LSCOM Lite and the rest of the 11 semantic concepts are selected from LSCOM by considering their availability and detectability.

Relevance Scores or *Confidence Scores* for the lexicon of semantic concepts are assigned to each keyframe. The relevance score indicates semantic weight ranging from -1 to +1, shows the relevance between a key-frame (i.e. shot) and a particular semantic concept, where -1 implies *highly irrelevant* and +1, *highly relevant*. IMARS detects a set of semantic concepts using *Support Vector Machines* (SVMs) which were trained using various low-level visual features such as color histogram, color correlogram, edge histogram, etc extracted at regional and global level. The relevance score shows the closeness of an image to the decision boundary of the Support Vector Machine trained for that semantic concept. This relevance score can be obtained using any supervised binary classifier.

Accuracy of a semantic concept detector often depends upon three things - low-level features used for training, diversity in training set images and machine learning classifier used. For constructing a semantic concept detector, IMARS extracts varieties of low-level color and texture global features from diverse training images, where each feature was extracted at the global and regional level and separately trained with SVM. Even though the local features such as SIFT, SURF, GLOH etc. perform much better than the global features for semantic concept detection [52], the local feature based semantic concept detection involves construction of Bag-of-Visual-Words (BoW) model which is computationally very expensive. Even though global features sometimes suffer from less discriminability, IMARS solves the performance gap of different global descriptors based SVMs by using ensemble fusion of individual SVM classification scores. Irrespective of the slow computation speed, probabilistic SVMs are mostly preferred for semantic concept detection because of its higher accuracy than other classifiers such as Naïve Bayes and Neural Networks [35]. Since performances of the binary semantic classifiers vary from each other, it is hard to determine a standard threshold value for binarizing the relevance scores into 0 or 1. So, instead of hard classification, the relevance scores are used as such (i.e. soft classification).

The vector of semantic weights (i.e. relevance scores), denoted as the *model vector* or *semantic multinomial*, is learned from the image content and not from metadata or relevance feedback information [36]. This model vector represents an image in a semantic space. IMARS provides a diverse set of semantic concept detectors for visual scene categories that covers *places*, *people*, *objects*, *settings*, *activities* and *events*. So the lexicon of semantic concepts used in this work is sufficient enough to represent a keyframe in a high dimensional semantic space.

Performance of the IMARS concept detection engine for 39 semantic concepts of the LSCOM lite lexicon on TRECVID benchmark is presented in [37]. IMARS achieves Mean Average Precisions (MAPs) between a range of 0.3126 and 0.3356 for 7 experimental runs, which comparatively outperforms several semantic concept detection frameworks. For IMARS's detection accuracy for individual semantic concepts in the lexicon used in this paper, the readers are recommended to refer IBM's technical reports on semantic concept detection in TRECVID. Occasionally, misclassifications happen in IMARS semantic concept detection, because the low-level color distribution of an image of a particular semantic concept sometimes resembles the low-level color distribution of another semantic concept. For example, images containing actors with colorful costumes which belong to the concept *people* are sometimes misclassified as the concept *flowers*.

Specific movie events such as *action*, *comedy*, *dance* and *romance* are manually annotated for each shot. These movie events commonly occur in most of the movies which cannot be automatically detected even using state-of-the-art semantic concept detection methodologies because of two reasons.

i. These movie events are highly subjective,

ii. The low-level visual features trained for classification are not highly correlated with the corresponding event.

However, it is possible to detect action events using motion features. Since high camera motions might result in lower detection accuracy, manual annotation is adopted for annotating action events. In manual annotation of these movie events, relevance scores take value either 0 or 1, where 0 implies that the shot is irrelevant to the corresponding movie event and 1 implies that it is relevant. In this work, rare movie events and rare semantic concepts are not adopted for high-level feature annotation. The main reasons are:

i. Rare high-level features increases sparsity in the model vector,

ii. Less availability of high-level features in a movie might fail to satisfy the users' interests when the rare events or rare semantic concepts are chosen as preferences.

Presence of characters (i.e. movie actors) in a movie is automatically identified by recognizing the actors' faces in a movie video. A face recognition system [38] implemented in OpenCV was used for indexing movie characters. It recognizes frontal faces using *Eigenfaces* [39] with promising recognition accuracy. The Eigenface approach achieves recognition accuracies 95.7% and 95.3% on YALE and FERET face recognition datasets correspondingly [40]. However, the Eigenface approach sometimes misclassifies face images with different illumination, pose and rotation effects. To handle this, our prior work in [41] is used for illumination and pose invariant face recognition. The face recognition approach in [41] achieves 93.2% recognition accuracy on IDES dataset that contains face images with different illumination and pose variations, where the Eigenface approach attains recognition accuracy only about 78.4% on the IDES dataset. Here, the face images were extracted from actors' images that are crawled from the web, to train the face recognizer.

A graph based methodology [42] was used for scene change detection with a minor change: instead of using color and motion features, relevance scores from semantic concept detection (i.e. model vector) are used as visual features for constructing the *Shot Similarity Graph* (SSG). Similarity between two shots is computed using the *Cosine Similarity* metric. The SSG is segmented into sub-graphs, which are scenes of the movie, using *Normalized Cut* graph partitioning [43]. There is no standard threshold value for controlling the scene change detection. A high threshold might under segment the SSG, thus will result in too few scenes. Also a low threshold might over segment the SSG which will result in too many scenes. So, the number of scenes to be detected from a movie is determined manually.

The purpose of video segmentation is to reduce the computational complexity by breaking a large size problem into multiple smaller size ones, while preserving the optimality of the solution. Therefore, even if the shot/scene boundaries are missed or misclassified, it will not affect the optimality of the solution [44].

Speech transcripts of each shot are manually annotated using the *subtitles* file of a movie. Subtitles contain spoken content as sentences along with their starting and ending timings in the corresponding movie video. Since a movie is segmented into a set of shots based on the visual properties of the video, spoken sentences in subtitles will not have the same starting and ending timings as the shots. As such, speech transcripts of each shot are manually synchronized with the spoken sentences in the subtitles.

In the context of movies, the audio of a movie is mostly utilized by the textual modality (i.e. speech transcripts). Accordingly, classification of audio segments into concepts such as *applause*, *cheering*, *music*, *speech*, etc. is not adopted in this work. Moreover, some of these widely used audio semantic concepts seldom occur in movies (*applause*, *cheering*), whilst some of them occur throughout the movie (*speech*, *music*). There is therefore no need to explicitly annotate these audio semantic concepts.

3.2. User Interface

Fig. 3 depicts the graphical user interface of the proposed movie summarization system. The source code of the system can be freely downloaded at https://sourceforge.net/projects/moviesummarizer/.

The user interface module is used for the construction of a user profile and movie summary presentation (as shown in fig. 2). The user interface allows a user to construct their profile with multiple preferences. The preferences for movie summarization are *characters*, *events*, *semantic concepts* and *keywords* which are semi-automatically annotated in pre-processing. Usually, a user profile would consist of *demographic data* and *preferences*. Demographic data will contain basic information about the user such as *age*, *gender*, *country*, etc. Preferences will contain set of *semantic tags* and their corresponding *weights*. The weight denotes the importance of a particular semantic tag to the corresponding user. This user profile can be considered as the one used in recommender systems. However, the proposed system maintains the user profiles only during the summarization.

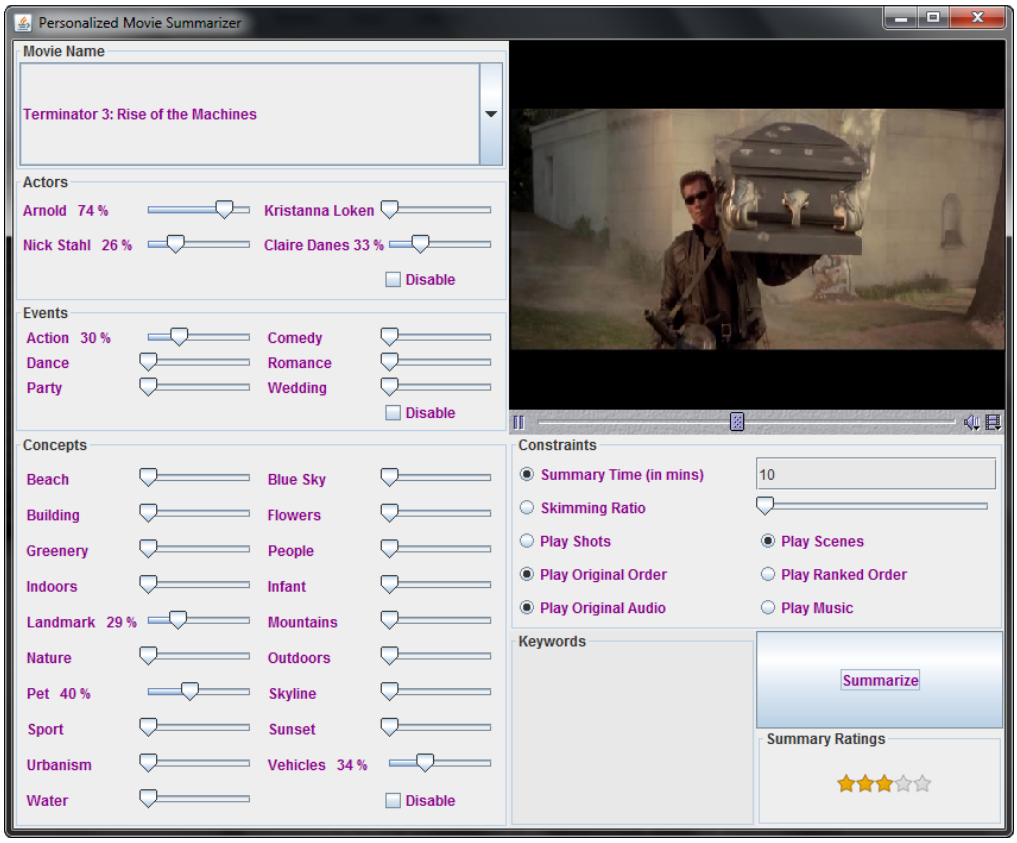

Fig. 3 User interface of the proposed movie summarization system. Here, users can select the preferences on actors, events, semantic concepts and keyword. Constraints for the summarization can be specified using the constraints panel. The video panel will render the video summary based on the user preferences and constraints

In general, preferences are collected from users either *explicitly* or *implicitly*. Explicit collection of preferences is achieved by directly asking users to specify a list of items liked by them, or to rate items using sliding scales. Preferences can be implicitly collected by indirectly observing previously purchased or viewed items of a user. Since implicit preference collection is not suitable here, explicit preference collection is adopted for personalized movie summarization. The interface of the proposed system makes it easier for the users to explicitly specify their preferences.

When a user selects a movie, actors' names shown in the user interface will be automatically updated to the actors of the currently selected movie. *Events* and *semantic concepts* shown in the user interface are common for all the movies, though. The *event panel* contains the four manually annotated movie events as well as two automatically detected semantic concepts namely *party* and *wedding*. Since, these two aforementioned semantic concepts are types of events; they are placed in the event panel for uniformity. The *concepts panel* contains rest of the 19 automatically detected semantic concepts. User profile is explicitly constructed from the user's input obtained from the query interface. Since binary valued preferences treat all the preferences equally, it cannot be used to comparatively prioritize the preferences. So, sliders are used to indicate the weight (i.e. importance) of a preference using real values from 0 to 1. Now the users can comparatively prioritize the preferences using the sliders. In the user interface, users can directly indicate their preferred keywords. Preferences,

which are the set of *characters*, *events*, *semantic concepts* and *keywords* preferred by a user, will be considered as user profile.

The *constraints panel* assists users to control the whole summarization process by specifying various constraints. The duration of the desired summary can be specified using either *summary time* (in minutes) or *skimming ratio* (in percentage). Users can also determine the type of the summarization, by selecting either summarization at shot level or summarization at scene level. The order of the selected shots or scenes for a summary skim is determined by choosing either *original order* or *ranked order*. If a user selects *original order*, the segments that are selected for the summary will be played by following their corresponding order in the movie. If a user selects *ranked order*, the segments that are highly relevant to the user's preferences will be played first. Sometimes, the audio transitions of segments in the summaries generated at shot level might be disrupting. So, to control audio of the summary, users can prefer to listen to either the original audio of summary or a music track from the corresponding movie [45]. Segments that are selected for the summary are concatenated and shown to the user in the *video panel*. The video panel will indicate the location of currently playing summary video segment in the movie. The likability of the summary shown is directly obtained from the users by employing a five star rating scale.

The proposed system is implemented in Java. The Java Media Framework is used to handle audio and video for summary presentation. The system uses flat files (CSV files) for maintaining metadata of movies (i.e. shot, scene segmentation data, semantic concepts, events and face annotation data, and speech transcripts). Since the metadata are maintained in a standard format, the proposed system can be used with any pre-processing tools where only a little effort is needed to convert the outcome of any pre-processing tool into the system supportable format.

3.3. Video Summarization

In the video summarization module, video segments (i.e. shots or scenes) are ranked based on the similarity calculated between the user profile and high-level movie features. Modified Dot Product, Jaccard coefficient and Jaccard set similarity are used for ranking individual video segments. The segments are selected for the summary skim based on ranking and the user's constraints on summarization. These summarization results are then sent to the user interface as depicted in fig 2. Selected segments are concatenated based on the user's requirements and played to the user using the user interface. The video summarization methodology will now be presented in more detail in the following section.

4. Summarization Methodology

Tailored movie summarization can be viewed as a process of measuring the similarity score of each video segment for the given user preferences and selecting those top ranked segments that will increase the cumulative similarity score of the summary. Here, summaries can be generated either at shot level or at scene level. Similarity scores for user preferences of *characters*, *events* and *semantic concepts* and *keywords* are measured for ranking each video segment. All these similarity scores are fused into a single *semantic similarity score* (σ). Fig. 4 depicts the work flow of the proposed summarization methodology.

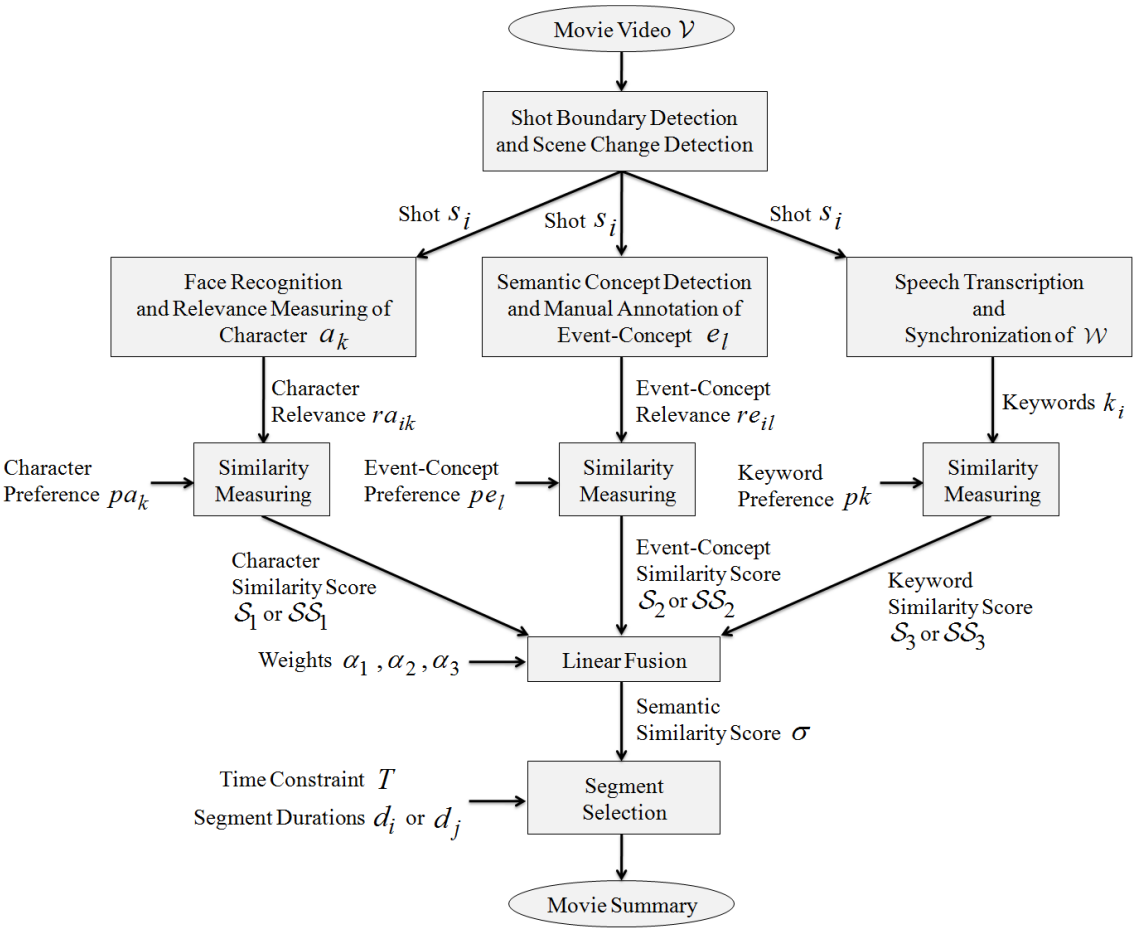

Fig. 4 Work flow of the proposed summarization methodology

The given movie V is segmented into a set of n shots $\mathrm{V} = \{s_1, s_2, ..., s_n\}$, where each shot s_i has a duration d_i seconds. Let the movie V consist of m scenes $\mathrm{V} = \{c_1, c_2, ..., c_m\}$, where each scene c_j = < collection of z shots >. Each scene c_j has a duration d_j seconds, and a shot in scene c_j is denoted by s_y. Let the set of leading *Characters* in V be $\mathrm{A} = \{a_1, a_2, ..., a_p\}$. Relevance of a character a_k in shot s_i is denoted by ra_{ik}, which is measured as,

$$ra_{ik} = \frac{\log f_{ik} - \min_{ik}(\log f_{ik})}{\max_{ik}(\log f_{ik}) - \min_{ik}(\log f_{ik})} \tag{1}$$

The term f_{ik} denotes the number of faces recognized in shot s_i for the character a_k using face recognition. The logarithm transforms f_{ik} into a smaller range. Since the frame rate of a video directly affect the number of faces recognized, ra_{ik} is normalized so that, it will take values between 0 and 1. Let pa_k denote the preference of the user on character a_k, where $pa_k \in \mathrm{R}$, $0 \le pa_k \le 1$.

Events and semantic concepts are considered as one group which will be denoted by *Event-Concept*. Let $\mathrm{E} = \{e_1, e_2, ..., e_q\}$ denote the set of *Events-Concepts* used for annotating V. Relevance of *Event-Concept* e_l for the shot s_i is denoted by re_{il}. For manually annotated *Events-Concepts*, re_{il} takes value either 0 or 1. In automatic semantic concept detection, re_{il} is the relevance score calculated from a semantic classifier for e_l. Since the relevance scores are between -1 and +1, and the system allows multiple preferences, a higher negative relevance score for a preferred semantic concept will reduce the similarity score, even though a shot has many higher positive scores for other preferred semantic concepts (*false negatives*). This will also increase the chances for the shots with lower negative relevance scores of semantic concepts to enter in the summary (*false positives*). As a solution, relevance scores in the range -1 to +1 can be normalized into the range 0 to 1. Since this *range normalization* is the linear transformation of values, negative relevance scores will take values between 0 and 0.5; positive relevance score will be between 0.5 and 1. So, the effect will still remain the same even after range normalization. In order to solve this problem, the negative relevance scores are ignored and are assumed to be zero. Now, re_{il} would take values from 0 to 1. Let pe_l denote the preference of the user on *Event-Concept* e_l, where $pe_l \in \mathrm{R}$, $0 \le pe_l \le 1$.

Let $\mathrm{W} = \{k_1, k_2, ..., k_n\}$ denote a set of *Keywords*, which are semi-automatically annotated speech transcripts, where each k_i = < collection of keywords of s_i >. Let pk denote a set of keywords preferred by the user.

4.1. Shot Level Summarization

The similarity between each shot and the user preferences on *Characters*, *Events-Concepts* and *Keywords* is measured. Numeric preferences are interpreted as quantity requirements; so any metric from the *inner product family* [46] can be used as a similarity metric. *Modified Dot Product* is used to measure the *character similarity score* between ra_{ik} and the preference pa_k for shot s_i, using a function S_1 defined as,

$$\mathrm{S}_1(s_i, ra_{ik}, pa_k) = \frac{1}{p} \sum_{k=1}^{p} ra_{ik} \cdot pa_k \tag{2}$$

where $\mathrm{S}_1 \in \mathrm{R}$, $0 \le \mathrm{S}_1 \le 1$. The *event-concept similarity score* for shot s_i over the preference pe_l is computed as a *Jaccard coefficient* using a function S_2 defined as,

$$\mathrm{S}_2(s_i, re_{il}, pe_l) = \frac{\sum_{l=1}^{q} re_{il} \cdot pe_l}{\sum_{l=1}^{q} re_{il}^2 + \sum_{l=1}^{q} pe_l^2 - \sum_{l=1}^{q} re_{il} \cdot pe_l} \tag{3}$$

where $S_2 \in R, 0 \leq S_2 \leq 1$. For each shot s_i, *Jaccard similarity* is used to measure the *keyword similarity score* between k_i and pk using a function S_3 defined as,

$$S_3(s_i, k_i, pk) = \frac{|k_i \cap |}{|k_i \cup |} \tag{4}$$

where $S_3 \in R, 0 \leq S_3 \leq 1$. The *semantic similarity score* $\sigma(s_i)$ for each shot s_i can be computed as a linear fusion of the similarity scores computed using functions S_1, S_2 and S_3,

$$\sigma(s_i) = \sum_t \alpha_t S_t \tag{5}$$

where $\alpha_1, \alpha_2, \alpha_3 \geq 0$ and $\alpha_1 + \alpha_2 + \alpha_3 = 1$. The α_1, α_2 and α_3 are weights used for linear fusion. It can be seen that the similarity scores computed using S_1, S_2 and S_3 are in the range of 0 and 1. Because a similarity score in the higher range will enjoy a higher priority among other similarity scores. Now only the weights α_1, α_2 and α_3 can determine the priority among different similarity scores.

4.2. Scene Level Summarization

For each scene c_j, the similarity scores of *Characters*, *Events-Concepts* and *Keywords* are computed using functions SS_1, SS_2 and SS_3 respectively. For the given preferences pa_k, pe_l and pk, these scores are measured as,

$$SS_1(c_j, ra_{yk}, pa_k) = \frac{1}{z} \sum_{y=1}^{z} S_1(s_y, ra_{yk}, pa_k) \tag{6.1}$$

$$SS_2(c_j, re_{yl}, pe_l) = \frac{1}{z} \sum_{y=1}^{z} S_2(s_y, re_{yl}, pe_l) \tag{6.2}$$

$$SS_3(c_j, k_y, pk) = \frac{1}{z} \sum_{y=1}^{z} S_3(s_y, k_y, pk) \tag{6.3}$$

where $SS_1, SS_2, SS_3 \in R, 0 \leq SS_1, SS_2, SS_3 \leq 1$.

The terms $S_1(s_y, ra_{yk}, pa_k)$, $S_2(s_y, re_{yl}, pe_l)$ and $S_3(s_y, k_y, pk)$ can be calculated using equations (2), (3) and (4) respectively. The term $1/z$ used in equations (6.1), (6.2) and (6.3) penalizes scenes with higher number of shots not to get benefit from higher similarity score. The semantic similarity score $\sigma(c_j)$ for each scene c_j can be computed as a linear fusion of the similarity scores computed using functions SS_1, SS_2 and SS_3,

$$\sigma(c_j) = \sum_t \alpha_t SS_t \tag{7}$$

Here α_1, α_2 and α_3 are the same weights as used in equation (5). Since these weights are used only for prioritizing the similarity scores of different movie semantics, same weights can be used in linear fusion for both shot level and scene level summarization. Depending upon the problem and its application, the weights for linear fusion can be user-defined, equal, unequal, time varying, etc. It is trivial to ask the users to set these weights manually. Since equal weights do not prioritize the similarity scores, expert-defined unequal weights are used here. Based on the importance of each similarity score, the weights were determined by three media production professionals. Since characters are the most significant high-level features of a movie, α_1 should be greater than α_2 and α_3. Also, α_2 should be greater than α_3, since events and semantic concepts are considered more important than spoken content. By considering these objectives, summaries generated for different preferences with different set of weights were analyzed for arriving at optimal weights. Based on a pilot study with 3 experts, the weights α_1, α_2 and α_3 were set to 0.43, 0.36 and 0.21 respectively for the experiments.

4.3. Segment Selection

The objective of segment selection is to select the video segments (shots or scenes) that maximize the cumulative semantic similarity score for the summary while not exceeding a user's given time constraint T. This can be considered as an instance of the 0-1 knapsack problem [47] which is a combinatorial optimization problem. Given a set of items, each with a weight and a value, the objective is to find a set of items which maximizes the total value without exceeding the given weight limit, where each item can be selected only once (either 0 or 1). Here, the set of items is the set of video segments, each with a duration and a semantic similarity score, and the weight limit is the time constraint T. With an objective to find a set of video segments to maximize the cumulative similarity score, each video segment is either selected (1) or rejected (0) for the summary skim. For shot level summarization it is defined as,

$$max \sum_{i=1}^{n} \sigma(s_i) \cdot x_i$$

$$subject\ to \sum_{i=1}^{n} d_i \cdot x_i \le T \tag{8}$$

For scene level summarization it is defined as,

$$max \sum_{j=1}^{m} \sigma(c_j) \cdot x_j$$
$$subject\ to \sum_{j=1}^{m} d_j \cdot x_j \le T \tag{9}$$

x_i is a binary decision variable that takes the value 1 if shot s_i is selected for summary, 0 otherwise (same for x_j in the case of equation (9)). This 0-1 knapsack problem can be solved using branch-and-bound or greedy algorithm [47]. The proposed system employs a greedy algorithm because of its faster execution speed. The video segments with higher semantic similarity score are selected in each iteration until no more video segments can be selected under the given time constraint. Selected segments are ordered either in original order or in ranked order based on the user's requirement. The summary is skimmed by concatenating the selected and ordered segments, and showed to the user by considering the corresponding preference for audio.

5. Experiments and Results

Subjective tests were conducted with users to evaluate the performance of the proposed system and summarization methodology. There is no standard procedure to evaluate performance of a video summarization technique. Evaluation of video summarization can be classified into *intrinsic* and *extrinsic* methods [48]. Using intrinsic evaluation method, the quality of a summary is evaluated by analyzing them directly. The criteria used are - judgment of fluency of the summary, coverage of key ideas of the source video, or similarity to an ideal summary generated by experts. This type of evaluation also uses a questionnaire to evaluate users' experience about the summaries. In extrinsic evaluation, a summary is evaluated with respect to its impact on the performance for a specific information retrieval task. One common extrinsic evaluation is the quiz method, where multiple choice questions derived from a video presentation are asked to the users both before and after watching the corresponding summary skim. The quality of summarization is then measured by the increase in quiz scores. This evaluation is mostly used for videos with uniformly informative content that cannot be divided into events (ex. documentary and educational videos). Hence, this approach might not be suitable for evaluating videos with diverse high-level features such as movies, consumer videos and sports videos.

Both evaluations have their disadvantages. In intrinsic evaluation, it is hard to derive an ideal summary. Even experts might not agree on which video segments to be included in the ideal summary. In extrinsic evaluation, users often find it difficult to distinguish different summarization methods. Also, it is hard to prepare the quiz questions in an objective manner. Since evaluation of video summarization is a highly subjective task, intrinsic evaluation is adopted here for evaluating the proposed movie summarization method. Therefore the subjective user evaluation is adopted to determine the quality of the produced summaries (tailored) by comparing them with ideal summaries (non-tailored) by conducting questionnaires. Also, the subjective evaluation on the usability of the proposed movie summarization system is assessed by conducting a questionnaire survey.

5.1. Material

Table 1 shows the eight movies in the English language used for the experiments on the subjective evaluation of the tailored and non-tailored summaries (presented in section 5.3), whilst Table 2 details the storyline for each of them. The storyline of each test movie presents an overall description about its content which helps to conceive the distribution of different semantic elements in a movie. Most of the movies chosen belong to the action/adventure genre. This was specifically done so since content with high dynamism/inter-frame variability is notoriously difficult to summarize.

Table 1 Detailed information about test movies

Movie ID	Movie Name	Genre	Length (in mins)	No. of Shots	No. of Scenes	No. of leading characters
1	Rush Hour (1998)	Action/ Comedy/ Thriller	98	1231	62	2
2	Hancock (2008)	Action/ Fantasy	92	1283	68	3
3	Quantum of Solace (2008)	Action/ Adventure/ Crime	106	2431	97	3
4	Terminator 3 - Rise of the Machines (2003)	Action/ Sci-Fi/ Thriller	109	1154	65	4
5	Bad Boys (1995)	Action/ Comedy/ Crime	118	959	61	3
6	Indiana Jones and the Kingdom of the Crystal Skull (2008)	Action/ Adventure	122	1243	72	4
7	Mission: Impossible III (2006)	Action/ Adventure/ Thriller	126	2823	92	3
8	Slumdog Millionaire (2008)	Drama/ Romance/ Thriller	120	1961	81	4

Table 2 Storylines of the test movies

Movie ID	Movie Name	Storyline
1	Rush Hour (1998)	Two cops, who are from different culture and background; who can't stand each other, eventually team up to save a kidnapped girl.
2	Hancock (2008)	A hard-living superhero, whose reckless actions routinely cost the city millions of dollars. He enters into a questionable relationship with the wife of a person who tries to save his public image.
3	Quantum of Solace (2008)	A secret agent tries to stop a wealthy business man who intends to seize control of a country's most valuable resource. At the same time, he still tries to seek revenge over the death of his love.
4	Terminator 3 - Rise of the Machines (2003)	A robotic warrior from future travels back in time to protect a 20-year old boy and his future wife. It saves them from a most advanced robotic assassin and to ensure they both survive a nuclear attack.
5	Bad Boys (1995)	Two detectives who protect a murder witness while investigating a case of stolen heroin.
6	Indiana Jones and the Kingdom of the Crystal Skull (2008)	A famed archaeologist is called back into action when he becomes entangled in a Soviet plot to uncover the secret behind mysterious artifacts known as the Crystal Skulls.
7	Mission: Impossible III (2006)	A secret agent who comes face to face with a dangerous and sadistic arms dealer while trying to keep his identity secret in order to protect his girlfriend.
8	Slumdog Millionaire	A teenager, who grew up in the slums, becomes a contestant of a reali-

(2008)	ty quiz show. He is arrested under suspicion of cheating, and while being interrogated, events from his life history are shown which explain how he knows the answers.

5.2. Participants

Twenty subjects (12 male and 8 female) participated in the subjective evaluation test. All participants were undergraduate and postgraduate students ranging in age from 18 to 23. They had no previous knowledge about video summarization or video editing. Participants self-reported that they watched television for an average of three and half hours a day. The main qualifying criterion for participating in the evaluation was for a participant to have previously watched all the test movies (Table 1), whose generated summaries were going to be evaluated. This, as it is reasonable to assume that in order for a participant to gauge the quality of a movie summary, s/he would have had to watch the full length movie in the first place.

Since evaluating summaries of all eight movies would have been burdensome for the subjects, they were randomly divided into 2 groups of 10 subjects each. Subjects of the first group evaluated the summaries of the first four test movies (as shown in table 1), whilst subjects in the second group evaluated summaries of the last four movies listed in table 1.

5.3. Subjective Evaluation of the Tailored and Non-Tailored Summaries

Subjects were first invited for a one hour tutorial session, in which they were given instructions about the system, but were not informed about the methodology used for summarization. In the session, they were shown how to specify preferences to summarize movies at both shot level and scene level for a desired video summary length. In order to see the effect of their preferences and familiarize themselves with the summarization software, subjects picked any movie of their choice from a selection of library DVDs (not the movies subsequently evaluated and detailed in Table 1), and were encouraged to vary their preference parameters in order to see the impact of the changes on the generated summaries.

Two days after attending the tutorial session, subjects were then invited to the first of two consecutive evaluation days. Subjects were instructed to create 10 minute (both shot and scene level) summaries for the movies they were about to abstract - since the test movies that were used in the evaluation have an average length of 110 minutes, this corresponded to roughly a 10% skimming rate. The summarization preferences for a movie were down to each individual subject; although these could change from movie to movie according to the subjects' tastes, for a particular movie, these were only specified once (at the outset) and a tailored summary generated as a result.

A generic or non-tailored summary of 10 minutes length for each movie was manually created at both shot level and scene level, which would contain the most interesting and significant content of a movie. To generate an ideal summary, first a movie is segmented at both shot level and scene level using the same shot boundary detection and scene change detection techniques which were used for tailored movie summarization. The total number of shots and scenes to be segmented for generic summary is also used as in table 1. Then, the most informative and interesting shots/scenes are manually selected for a total length of 10 minutes which are concatenated in original movie order with original movie audio. The resulting summaries are treated as shot level and scene level generic summaries of the test movies.

Once summaries had been generated for a particular movie, subjects were shown both shot level and scene level tailored and non-tailored summaries of the movie in question - in a ran-

domized order. Both tailored and non-tailored movie summaries are played to each subject in our user interface presented in fig 3. To indicate the distinction between summaries, starting and ending of each summary was clearly indicated using pop-up messages in the user interface. Subjects were allowed to watch both tailored and non-tailored summaries as many times as they wished. However, they were not told which summary was tailored and which wasn't, i.e. the evaluation was blind.

On each evaluation day, each subject assessed two movies (hence the need for two evaluation days). On average, it took subjects 55 minutes to evaluate a test movie. Once the evaluation of a movie was done, subjects were given a 10 minute break before the next movie was evaluated.

A questionnaire was prepared to evaluate tailored and non-tailored summaries. The questions used in the questionnaire were,

- How informative was the summary?
- How enjoyable was the summary?
- Is this summary relevant to your interests?
- How willing would you be to accept this summary?

These questions evaluate the summarization performance measures *informativeness*, *enjoyability*, *relevance* and *acceptance* respectively. For each measure, subjects were asked to rate both tailored and non-tailored summaries on a scale of 1-100 with 1 being very bad and 100 being very good. Table 3 depicts the comparative results of tailored and non-tailored summaries.

Table 3 Average scores on the summarization performance measures given by the subjects for Tailored summaries (T) and Non-Tailored summaries (NT) at shot level and scene level. Overall Average (Avg) and Standard Deviation (Stdv)

Movie ID		Informativeness		Enjoyability		Relevance		Acceptance	
		Shot Level	Scene Level	Shot Level	Scene Level	Shot Level	Scene Level	Shot Level	Scene Level
1	T	**67.2**	**83.3**	**59.4**	76.5	**63.4**	**83.2**	**74.6**	**81.5**
	NT	60.7	72.5	51.1	**89.3**	59.4	72.8	63.9	70.9
2	T	57.9	**84.6**	47.2	**81.7**	**72.3**	**89**	56.2	**83.5**
	NT	**61.3**	63.2	**52.9**	74	54.9	69.3	**60.3**	75.3
3	T	53.2	75.4	**66.7**	**82.2**	**84.8**	77.4	**79**	**83.4**
	NT	**61.6**	**73.8**	59.9	77.1	63.5	**81.9**	68.2	81.3
4	T	**63.7**	**82**	**49.2**	**84.9**	**81**	**90.3**	**64.6**	**79.2**
	NT	58.2	72.3	48.4	73.5	79.8	82.4	53.1	72.4
5	T	**52.5**	**78.9**	53	**88.2**	69.3	**79.5**	**55.3**	**75.3**
	NT	51.4	69	**57.2**	79.5	**70.2**	75.5	48	71.8
6	T	**72.3**	77.6	**65.9**	**87**	**79.3**	**89.2**	**64**	**84.4**
	NT	68.8	**86.4**	52.4	74.2	64.2	68.3	63.2	71.9
7	T	**54.7**	**80.5**	**53.2**	**83.6**	**72**	**72.4**	**56.3**	78.1
	NT	53	74.1	50	80.1	63.8	63.3	47.9	**82.4**
8	T	**64.8**	67	**59.7**	**74.9**	57.3	**75**	**60.3**	80
	NT	56.3	**86.5**	55.2	72.4	**60.6**	72.9	53.5	**87.8**
Avg	T	60.7	78.6	56.8	82.3	72.4	82	63.7	80.6
	NT	58.9	74.7	53.3	77.5	64.5	73.5	57.2	76.7
Stdv	T	10.3	9.8	10.2	8.6	11.8	10.5	11.9	8.0
	NT	8.8	10.3	8.4	8.6	10.0	8.9	10.1	10.2

A number of conclusions can be reached by observing table 3. Summaries at scene level are more informative than those at shot level. Since shots have smaller durations, they cannot convey much information coherently. So both tailored and non-tailored summaries at shot level perform evenly in terms of them being informative. However, sometimes non-tailored summaries at scene level are more informative than tailored summaries at scene level. The reason is that non-tailored summaries are created manually by selecting the most informative and exciting scenes of movies. Also, summaries at scene level are more enjoyable than summaries at shot level, because audio-visual transitions between shots might be disrupting sometimes. So both tailored and non-tailored summaries at shot level perform closely similar in enjoyability measure. Since the system tailors the summary to the subject's preference, tailored summaries at both levels achieve better relevancy than the non-tailored summaries at both levels. It can be seen that the tailored summaries are widely accepted among subjects than the non-tailored summaries.

Performance of the summarization system can be assessed by two *Quality of Perception* measures, which are *Quality of Perception - Information Assimilation* (*QoP-IA*) and *Quality of Perception - Satisfaction* (*QoP-S*) [49]. QoP-IA denotes the user's ability to assimilate information from a multimedia presentation, and QoP-S implies the user's satisfaction from a multimedia presentation. QoP-IA is measured by averaging the informativeness scores and relevance scores of summaries (Equation 10). Because, information is assimilated only if the content shown to the user is relevant and informative. QoP-S is calculated by averaging the scores of enjoyability and acceptance (Equation 11).

$$QoP-IA=\frac{1}{2}\left(Informativeness+Relevance\right) \tag{10}$$

$$QoP-S=\frac{1}{2}\left(Enjoyabili\,ty+Acceptance\right) \tag{11}$$

Fig. 5 illustrates the evaluation results on quality of perception measure QoP-IA for tailored and non-tailored summaries at shot level and scene level for the test movies. It shows that the tailored summarization at scene level performs better than non-tailored summaries, since the former shows information relevant to each individual subject's interests. Also, when it comes to relevancy, non-tailored summaries do a relatively poor job of meeting the users' interests.

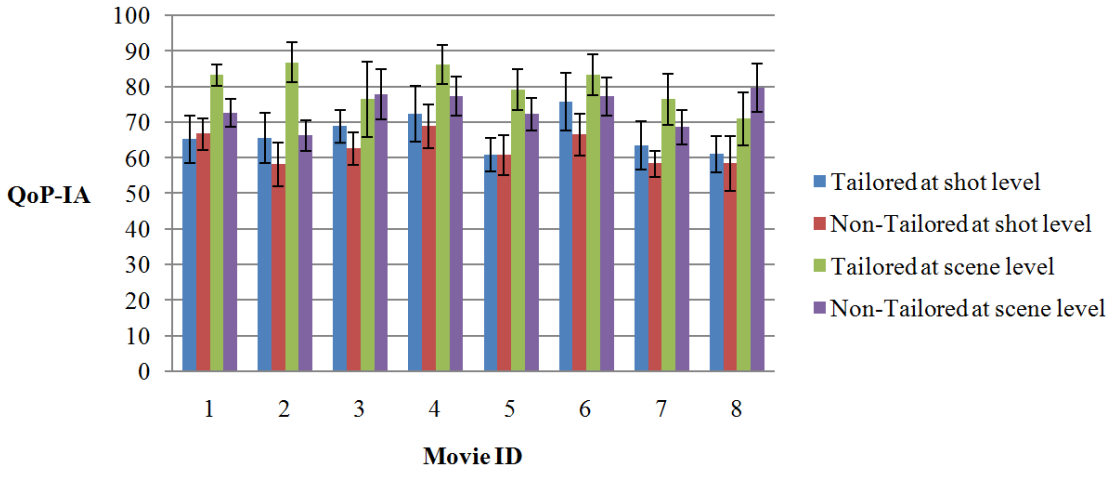

Fig. 5 Subjects' information assimilation from Tailored and Non-Tailored summaries of test movies at shot level and scene level

Evaluation results on the measure QoP-S for tailored and non-tailored summaries at shot level and scene level for the test movies are shown in Fig. 6. The results for the QoP-S measure show that the tailored summaries at scene level have higher scores than their non-tailored counterparts. It can be noticed that the personalization improves both enjoyability and acceptance of the summary. In both quality of perception measures, scene level summarization (both tailored and non-tailored) is more informative and satisfiable than shot level summarization (both tailored and non-tailored).

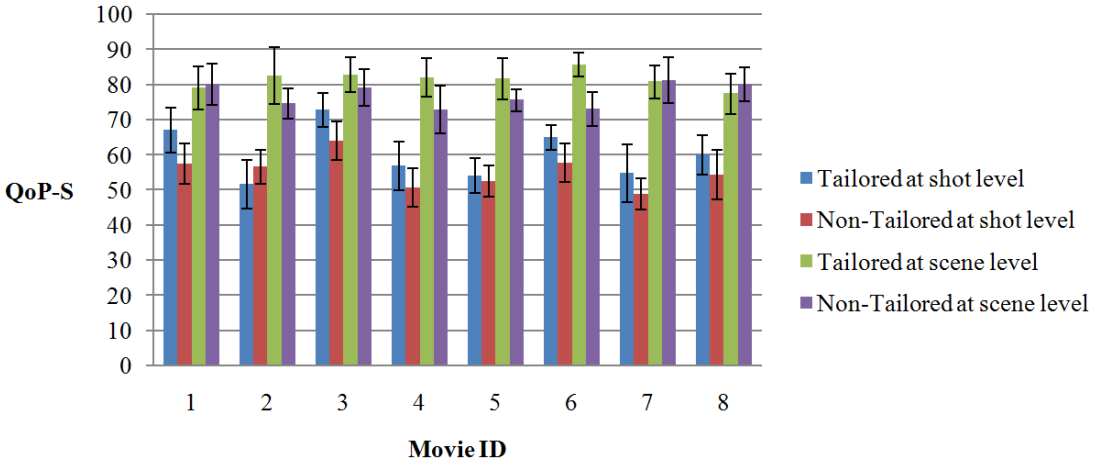

Fig. 6 Subjects' satisfaction on Tailored and Non-Tailored summaries of test movies at shot level and scene level

Statistical significance of the difference in the average scores of QoP-IA and QoP-S for the summarization techniques was assessed using *dependent t-test for paired samples*. Our null hypothesis is that summarization techniques do not affect the users' ability to assimilate information from a summary (QoP-IA) and satisfaction over the summary content (QoP-S). Table 4 and table 5 depict the *t* values calculated for the QoP-IA and QoP-S scores respectively. Variable 1 and variable 2 denote the QoP scores corresponding to the two summarization techniques which are going to be compared in a t-test. The tables show that the differences in their means are statistically significant, irrespective of whether the tailoring happens at scene level or shot level.

Table 4 Results of t-tests conducted for QoP-IA scores given for the four summarization techniques. V1 – Variable 1, V2 – Variable 2, μ (V1) – Mean of Variable 1, μ (V2) – Mean of Variable 2, σ^2 (V1) – Variance of Variable 1, σ^2 (V2) – Variance of Variable 2

V1	V2	μ (V1)	μ (V1)	σ^2 (V1)	σ^2 (V2)	T test
Tailored at shot level	Non-Tailored at shot level	66.60	61.73	63.57	43.70	t=5.125; p<.05
Tailored at scene level	Non-Tailored at scene level	80.33	74.13	67.29	46.80	t=4.753; p<.05

Table 5 Results of t-tests conducted for QoP-S scores given for the four summarization techniques. V1 – Variable 1, V2 – Variable 2, μ (V1) – Mean of Variable 1, μ (V2) – Mean of Variable 2, σ^2 (V1) – Variance of Variable 1, σ^2 (V2) – Variance of Variable 2

V1	V2	μ (V1)	μ (V1)	σ^2 (V1)	σ^2 (V2)	T test
Tailored at shot level	Non-Tailored at shot level	60.29	55.31	77.13	44.80	t=5.210; p<.05
Tailored at scene level	Non-Tailored at scene level	81.52	77.11	34.91	35.62	t=4.326; p<.05

5.4. Subjective Evaluation on System Usability

The same 20 participants were invited for a one day system usability evaluation experiment after the subjective evaluation of tailored and non-tailored summaries. To effectively evaluate the usability of the proposed system, each participant was separately provided with the proposed summarization software and the library DVDs. The participants were then asked to use the system for generating personalized movie summaries. The choice of movies to be summarized using the system was left to the individual subject. Even though subjects were not given any time restriction regarding summarization system usage, they all ended up using it between 60 to 90 minutes to generate their summaries.

Table 6 Result of subjective evaluation on summarization system usability

ID	Statement	Number of subjects rated on 1-5 scale					Average Rating
		1	2	3	4	5	
S1	I am willing to prefer Actors as my preferences	1	3	5	8	3	3.45
S2	I am willing to prefer Events as my preferences	0	3	2	12	3	3.75
S3	I am not willing to prefer Concepts as my preferences	6	1	8	4	1	2.65
S4	I am not willing to prefer Keywords as my preferences	4	0	8	4	4	3.20
S5	The system generate summaries that match my preferences	0	1	3	7	9	4.20
S6	I don't like summaries generated at shot level	2	9	4	4	1	2.65
S7	I like summaries generated at scene level	0	0	3	12	5	4.10
S8	I don't like summaries generated in original order	14	4	1	1	0	1.45
S9	I don't like summaries generated in ranked order	1	2	3	7	7	3.85
S10	I like summaries generated with original audio	1	0	1	7	11	4.35
S11	I like summaries generated with musical audio	4	1	4	3	8	3.50
S12	The system doesn't help me to understand a movie	10	5	3	2	0	1.85

Subjects completed a questionnaire (as shown in Table 6) evaluating the usability of the proposed summarization system. The subjects were then asked to rate each statement on a Likert Scale of 1-5 with 1 being strong disagreement and 5 being strong agreement. To prevent bias, the questionnaire included an equal mix of positively and negatively phrased

statements, randomly spread throughout the questionnaire. Moreover, the *Think-aloud protocol* method was also adopted to gather the subjects' verbal opinions about usability of the summarization system.

Table 6 shows the result of the subjective evaluation on the summarization system usability. It includes questionnaire for the system usability evaluation, the number of subjects who have given corresponding ratings and the average ratings.

The statements used in the usability evaluation can be classified into one of the two types. The first type of statements (S1 to S4, and S6 to S11) considers subjects' opinion about the usability of the features provided by the proposed summarization system. The second type of statements (S5 and S12) assesses the overall performance of the proposed system in terms of achieving personalization (S5) and summarization (S12).

Fig. 7 depicts the number of statements rated on a 1-5 Likert Scale by each subject. It shows the subjects' diverse ratings to the evaluation statements, where most of the subjects utilized the Likert Scale range 1-5 to express their opinions on the summarization system usability. For ease of interpretation and uniformity, responses to negatively phrased statements (S3, S4, S6, S8, S9, and S12) have been positively coded. Fig. 7 shows that the proposed system is accepted by each subject, and performs fairly under all the usability evaluation statements.

The statements S1 to S4 record the subjects' opinion about each type of preferences (i.e. *actors*, *events*, *concepts* and *keywords*) provided in the proposed system. While interacting with the subjects, it was observed that the subjects have different opinions about the preferences supported by the system.

Regarding S1, the subjects preferred actors as preferences most of the time. But some subjects stated that they gave less importance to the actors to summarize a movie, and they reported that they would give importance to movie events instead. The subjects' choice among actors varied each time they wanted to summarize a movie with different preferences. It was noticed that the choice of actors sometimes depends upon the subjects' familiarity with the movie actors. Some subjects reported that they barely prefer unfamiliar actors for movies which they have not watched yet. In that case, they stated that they were interested in other common preferences such as events and concepts. However, most of the other subjects reported that they prefer the leading actors irrespective of their familiarity with the corresponding movie actors.

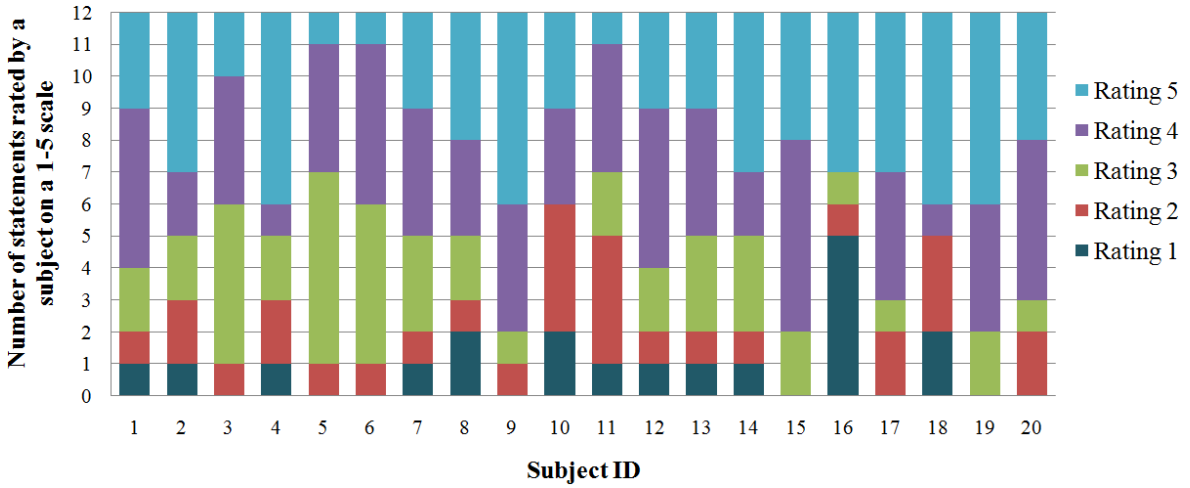

Fig. 7 The number of statements rated on a scale of 1-5 by each subject

From the subjects' responses to S2, it can be inferred that most of the subjects prefer events as their preferences to summarize a movie. It was found that only a very few subjects did not prefer events as preferences. The choice of events however varied among each individual subject and changed each time they wanted to summarize a movie. Among all the events provided in the summarization system, events such as *party* and *wedding* were not preferred by the subjects frequently. Some of the subjects suggested us to include specific action events such as *explosion* and *gunshot* to the system in the future.

Considering the subjects' responses to S3 and S4, it was observed that most of the subjects awarded less importance for the semantic concepts and keywords than actors and events. It was noticed that most commonly occurring semantic concepts such as *indoor* and *outdoor* were seldom preferred to summarize a movie. Also, few subjects were interested in the rare semantic concepts such as *sports*, *pet*, *vehicle* and others. Some subjects were highly interested to choose keywords as their summary preference. They reported that it was amusing to summarize movies with specific choice of keywords. Moreover, some subjects responded that they are more interested in visual content than the spoken words.

During the evaluation, subjects chose multiple preferences with different weights most of the time. Statement S5 evaluates the summarization system's ability to generate a personalized movie summary by matching user's preferences against the movie content. From the average ratings, we can conclude that the system generates personalized movie summaries that match subjects' interests. However, a few subjects reported that the system sometimes did not include some preferred movie features in the summary. This sometimes happens because of the less availability of some preferred movie events, concepts and keywords in a particular movie.

Regarding the summary presentation, statements S6 to S11 evaluate the subjects' impression about the features that directly affect the way tailored summaries are presented to the users in terms of the summary segments' granularity (*shots* vs. *scenes*), order (*original* vs. *ranked*) and audio (*original* vs. *musical*).

The average ratings for S6 and S7 denote that most of the subjects like scene level summaries than shot level summaries. Nonetheless, the subjects came up with different impressions about the granularity of the summary content, because each type of summaries has its own advantages. Most of the subjects argued that scene level summaries are highly informative and easy to understand than the shot level summaries. Because the shot level summaries often have faster audio-visual transitions than scene level summaries, scene level summaries are smoother and more informative than shot level summaries. This happens because the proposed system performs shot boundary detection only using visual features, sometimes the shots with very smaller durations often tend to contain a discontinuous audio segment. Thus, the important speech or musical information contained in the audio segments is divided into multiple video shots. So, a more informative shot level summarization can be achieved by using a combination of the audio-visual features for determining the shot boundaries in a movie. A few subjects supported the shot level summaries by stating that the shot level summaries cover a wide variety of movie elements within a smaller time period. They also stated that, sometimes the scene level summaries contain very lengthy scenes with a - limited number of interesting elements in it. A solution to this problem can be achieved by adopting over-segmentation of Shot Similarity Graph in scene change detection. That is, a slightly lower threshold for Normalized Cut graph partitioning will result in a higher number of scenes with a smaller number of shots. However, the number of scenes in a movie should be carefully determined.

Results for S8 and S9 show that the summaries generated at the original order are desired by the subjects most of the time. Many subjects responded that the summaries in the original

order help them to understand the flow of the movie content since it preserves the story order. However, fewer subjects reported that they liked summaries in the ranked order. Somewhat strangely, some subjects stated that they did not find any difference between the summaries generated in the original order and in the ranked order.

The subjects' responses to S10 and S11 denote that the original audio of the movie content is mostly preferred by the subjects, as the subjects felt that the original audio of the summary conveys more information about the movie content. It was also discovered that the choice of an audio for the summarization sometimes depends upon the granularity of the summary elements (i.e. *shots* or *scenes*). Some subjects responded that the musical audio is more enjoyable than the original audio for the shot level movie summaries. However, for the scene level summaries, the original audio was mostly preferred than the musical audio.

Hence, in conclusion, the choice of summary presentation in terms of the summary segments' granularity (*shots* vs. *scenes*), order (*original* vs. *ranked*) and audio (*original* vs. *musical*) completely depends upon the individual user's interests, and therefore it should be left to the individual user.

To efficiently evaluate S12, the subjects were advised to use the system with the movies that they already watched. From the results of S12, it can be observed that almost all the subjects denied the statement S12, and agreed the proposed system as a tool for movie summarization. Even though the system does not consider any objective criterion for summarizing a movie, still it includes the informative content which helps the users to understand a movie. Some of the subjects stated that the proposed system can also be used as a tool for browsing or searching a movie's semantic elements.

Further, the *Computer System Usability Questionnaire* (CSUQ) [50] was used to assess the user interface usability of the proposed system. The statements used from CSUQ are,

- The interface of this system is pleasant.
- I feel comfortable using this system.
- It was simple to use this system.
- It is easy to find the information I needed.
- Overall, I am satisfied with how easy it is to use this system.
- Overall, I am satisfied with this system.

These statements evaluate the user interface usability under criteria *appearance*, *comfortability*, *simplicity*, *user friendliness*, *usability* and *overall performance* respectively. The subjects were then given the questionnaire, and were asked to rate each usability criterion on a Likert Scale of 1-7 (as suggested in [50]) with 1 being strong disagreement and 7 being strong agreement.

Fig. 8 shows the box plot of the ratings given by the subjects on the user interface usability. Each box denotes the distribution of ratings given by the subjects for a certain usability criterion. The system performs fairly under all the user interface usability criteria. The results show that the interface is simple, comfortable and friendly to the subjects. This also shows that most of the subjects were satisfied by the usability and overall performance of the user interface of the proposed summarization system.

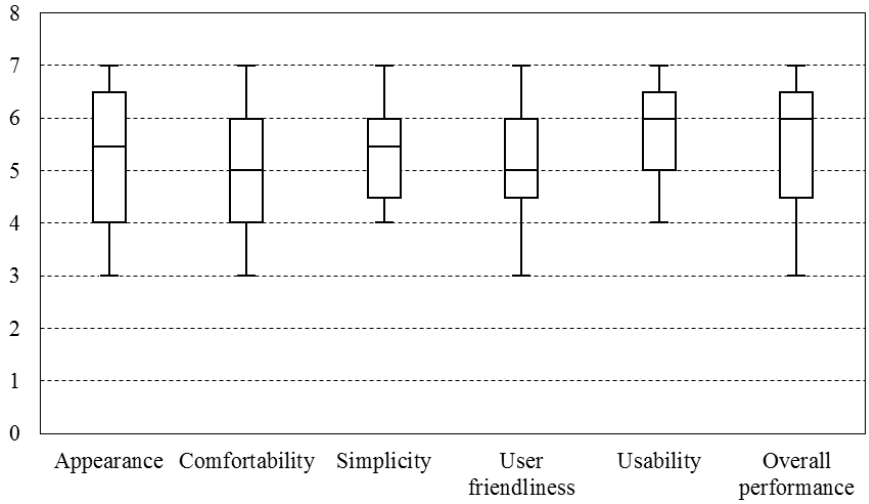

Fig. 8 Subjects' opinion on user interface usability

5.5. Limitations

The proposed movie summarization system also has some limitations. Subjective evaluation with a larger sample of users would have been nice. The subjective evaluation of tailored and non-tailored summaries was restricted to only action/adventure movies. Further, a few subjects argued that even though tailored summaries generated by the system match their interests, they do not sometimes include exciting and informative segments. Obviously, it is a tradeoff between whether to have a summary with informative segments and a summary with relevant video segments. Furthermore, as the proposed system does not consider any objective summarization criterion for summarization, it sometimes fails to include informative video segments in the summary. However, the system still randomly includes exciting and informative video segments into the tailored summaries.

6. Conclusion and Future Work

This paper presented a novel preference aware movie summarization system that produces semantically meaningful personalized movie summaries by adapting to the user's interest. As opposed to previous work which employed objective summarization criteria such as user attention, users' responding behavior, character analysis, information coverage, and diversity in order to accomplish generic video summarization, in our approach, personalized movie summarization is achieved by using users' subjective preferences as summarization criterion. The personalized summarization is attained using a unified approach comprising inner product similarity measures, linear fusion and a constrained selection scheme.

A detailed user study assessing the performance of the proposed summarization approach and the system usability are presented in this paper. Experimental results on personalized movie summarization against generic movie summarization demonstrate the effectiveness of the proposed system and the need for personalized summarization. The results of the subjective user studies on summarization have shown that movie summarization at scene level is more informative and satisfactory than summarization at shot level. Subjective evaluation on usability of the system demonstrated the users' diverse opinions on the proposed movie

summarization system. The results have also shown the potential usability of the proposed summarization system. Thus the results of the subjective user study have proved our hypothesis that is; a single generic movie summary does not satisfy every user where movie summaries should be generated based on individual user's preferences.

In future, performance of the proposed system will be assessed with movies from more diverse range of genres. Also, experiments will be conducted with subjects belong to different age groups. Though the system produces summaries which satisfy the users, tailored summaries can also be supplied with highly informative content. In future, the system would aim at providing relevant as well as informative summaries using a combination of subjective and objective summarization criteria. Also, performances of different fusion schemes other than linear fusion will be analyzed with the system.

Personalized summarization system resembles two other methods namely recommender system and relevance feedback system. Recommender systems and relevance feedback systems focus retrieval issues similar to personalized summarization. However, these two systems retrieve relevant content for the users, but do not summarize the content. In future, we think it will be interesting to explore how summarization systems will work in conjunction with recommender systems and relevance feedback systems for summarization.

References

1. Monaco J (2000) How to read a film: the world of movies, media, multimedia: language, history, theory. Oxford University Press USA.
2. Lu S, Lyu MR, King I (2005) Semantic video summarization using mutual reinforcement principle and shot arrangement patterns. In Proceedings of the 11th International on Multimedia Modelling 60-67.
3. Li B, Errico JH, Pan H, Sezan I (2004) Bridging the semantic gap in sports video retrieval and summarization. In Journal of Visual Communication and Image Representation, 15(3):393-424.
4. Smeaton AF, Over P, Doherty AR (2010) Video shot boundary detection: Seven years of TRECVid activity. Computer Vision and Image Understanding 114(4):411-418.
5. Sidiropoulos P, Mezaris V, Kompatsiaris I, Meinedo H, Bugalho M, Trancoso I (2011) Temporal video segmentation to scenes using high-level audiovisual features. IEEE Transactions on Circuits and Systems for Video Technology 21(8):1163-1177.
6. Money AG, Agius H (2008) Video summarisation: A conceptual framework and survey of the state of the art. Journal of Visual Communication and Image Representation 19(2):121-143.
7. Ma YF, Lu L, Zhang HJ, Li M (2002) A user attention model for video summarization. In Proceedings of the tenth ACM international conference on Multimedia 533-542.
8. Evangelopoulos G, Zlatintsi A, Potamianos A, Maragos P, Rapantzikos K, Skoumas G, Avrithis Y (2013) Multimodal saliency and fusion for movie summarization based on aural, visual, and textual attention. IEEE Transactions on Multimedia 15(7):1553-1568.
9. You J, Liu G, Sun L, Li H (2007) A multiple visual models based perceptive analysis framework for multilevel video summarization. IEEE Transactions on Circuits and Systems for Video Technology 17(3):273-285.
10. Ngo CW, Ma YF, Zhang HJ (2005) Video summarization and scene detection by graph modeling, IEEE Transactions on Circuits and Systems for Video Technology 15(2):296-305.
11. Taskiran CM, Pizlo Z, Amir A, Ponceleon D, Delp EJ (2006) Automated video program summarization using speech transcripts. IEEE Transactions on Multimedia 8(4):775-791.
12. Peng WT, Chu WT, Chang CH, Chou CN, Huang WJ, Chang WY, Hung YP (2011) Editing by viewing: automatic home video summarization by viewing behavior analysis. IEEE Transactions on Multimedia 13(3):539-550.

13. Joho H, Jose JM, Valenti R, Sebe N (2009) Exploiting facial expressions for affective video summarisation. In Proceedings of the ACM International Conference on Image and Video Retrieval. Article No 31.
14. Katti H, Yadati K, Kankanhalli M, Tat-Seng C (2011) Affective video summarization and story board generation using pupillary dilation and eye gaze. In IEEE International Symposium on Multimedia 319-326.
15. Tseng BL, Lin CY, Smith JR (2002) Video summarization and personalization for pervasive mobile devices. In Electronic Imaging 2002. International Society for Optics and Photonics 359-370.
16. Parshin V, Chen L (2004) Video summarization based on user-defined constraints and preferences. In Proceedings of RIAO 18-24.
17. Park HS, Cho SB (2011) A personalized summarization of video life-logs from an indoor multi-camera system using a fuzzy rule-based system with domain knowledge. Information Systems 36(8):1124-1134.
18. Lie WN, Hsu KC (2008) Video summarization based on semantic feature analysis and user preference. In Proceedings of IEEE International Conference on Sensor Networks, Ubiquitous and Trustworthy Computing 486-491.
19. Ghinea G, Kannan R, Swaminathan S, Kannaiyan S (2014) A Novel User-Centered Design for Personalized Video Summarization. IEEE International Conference on Multimedia and Expo (ICME) Workshop on Information Systems and Management in Multimedia Art, Education, Entertainment and Culture (MIS-MEDIA) 1-6.
20. Ghinea G, Thomas JP (1999) An approach towards mapping quality of perception to quality of service in multimedia communications. IEEE 3rd Workshop on Multimedia Signal Processing, 497-502.
21. Pfeiffer S, Lienhart R, Fischer S, Effelsberg W (1996) Abstracting digital movies automatically. Journal of Visual Communication and Image Representation 7(4):345-353.
22. Hermes T, Schultz C (2006) Automatic generation of Hollywood-like movie trailers. In cat1.netzspannung.org.
23. Li Y, Lee SH, Yeh CH, Kuo CC (2006) Techniques for movie content analysis and skimming: tutorial and overview on video abstraction techniques. IEEE Signal Processing Magazine 23(2):79-89.
24. Ellouze M, Boujemaa N, Alimi AM (2010) IM(S)^2: Interactive movie summarization system. Journal of Visual Communication and Image Representation 21(4):283-294.
25. Tobita H (2010) DigestManga: interactive movie summarizing through comic visualization. In Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems 3751-3756.
26. Shroff N, Turaga P, Chellappa R (2010) Video précis: Highlighting diverse aspects of videos. IEEE Transactions on Multimedia 12(8):853-868.
27. Fu Y, Guo Y, Zhu Y, Liu F, Song C, Zhou ZH (2010) Multi-view video summarization. IEEE Transactions on Multimedia 12(7):717-729.
28. Weng CY, Chu WT, Wu JL (2009) RoleNet: Movie analysis from the perspective of social networks. In IEEE Transactions on Multimedia, 11(2):256-271.
29. Tsai CM, Kang LW, Lin CW, Lin W (2013) Scene-based movie summarization via Role-Community Networks. IEEE Transactions on Circuits and Systems for Video Technology, 23(11):1927-1940.
30. Sang J, Xu C (2010) Character-based movie summarization. In Proceedings of the international conference on Multimedia 855-858.
31. Natsev A, Smith JR, Tešić J, Xie L, Yan R (2008) IBM multimedia analysis and retrieval system. In Proceedings of ACM international conference on Content-based image and video retrieval (CIVR) 553-554.
32. Kennedy L, Hauptmann A (2006) LSCOM lexicon definitions and annotations (version 1.0). DTO Challenge Workshop on Large Scale Concept Ontology for Multimedia, Columbia University ADVENT Technical Report #217-2006-3.

33. Naphade MR, Kennedy L, Kender JR, Chang SF, Smith JR, Over P, Hauptmann A (2005) A Light Scale Concept Ontology for Multimedia Understanding for TRECVID 2005. IBM Computer Science Technical Report RC23612 W0505-104.
34. Lu Y, Zhang L, Tian Q, Ma WY (2008) What are the high-level concepts with small semantic gaps? In IEEE Conference on Computer Vision and Pattern Recognition 1-8.
35. Chao C (2012) Subspace-based Semantic Concept Detection and Retrieval for Multimedia Information Systems. Open Access Dissertations, Paper 833.
36. Rasiwasia N, Moreno PJ, Vasconcelos N (2007) Bridging the gap: Query by semantic example. IEEE Transactions on Multimedia 9(5):923-938.
37. Amir A, Argillander J, Campbell M, Haubold A, Iyengar G, Ebadollahi S, Kang F, Naphade MR, Natsev AP, Smith JR, Tesic J, Volkmer T (2005) IBM research TRECVID-2005 video retrieval system. In Proceedings of the NIST TRECVID Workshop, Gaithersburg, MD.
38. Face Recognition using Eigenfaces, http://www.shervinemami.info/faceRecognition.html, Accessed 12 February 2013.
39. Turk M, Pentland A (1991) Eigenfaces for recognition. Journal of cognitive neuroscience 3(1):71-86.
40. Ruiz-del-Solar J, Navarrete P (2005) Eigenspace-based face recognition: a comparative study of different approaches. In IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 35(3):315-325.
41. Balakrishnan R, Kannan R, Kannaiyan S, Swaminathan S (2013) Deity face recognition using schur decomposition and hausdorff distance measure. IEEE 56th International Midwest Symposium on Circuits and Systems 1184-1187.
42. Rasheed Z, Shah M (2003) A graph theoretic approach for scene detection in produced videos. In Multimedia Information Retrieval Workshop.
43. Shi J, Malik J (2000) Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8):888-905.
44. Li Z, Schuster GM, Katsaggelos AK (2005) MINMAX Optimal Video Summarization, IEEE Transactions on Circuits and Systems for Video Technology 15(10):1245-1256.
45. Ghinea G, Chen SY (2006) Perceived quality of multimedia educational content: A cognitive style approach. Multimedia systems, 11(3): 271-279.
46. Cha SH (2007) Comprehensive survey on distance/similarity measures between probability density functions. International Journal of Mathematical Models and Methods in Applied Sciences 1(4):300-307.
47. Martello S, Toth P (1990) Knapsack problems: algorithms and computer implementations. John Wiley & Sons, Inc.
48. Taskiran CM (2006) Evaluation of automatic video summarization systems. In SPIE Conference Multimedia Content Analysis, Management and Retrieval 6073:178-187.
49. Gulliver SR, Ghinea G (2006) Defining user perception of distributed multimedia quality. ACM Transactions on Multimedia Computing, Communications, and Applications 2(4):241-257.
50. Lewis JR (1995) IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. International Journal of Human-Computer Interaction 7(1):57-78.
51. Zhang L, Xia Y, Mao K, Ma H, Shan Z (2014) An Effective Video Summarization Framework Toward Handheld Devices, IEEE Transactions on Industrial Electronics (In Press).
52. Le DD, Satoh SI (2011) A Comprehensive Study of Feature Representations for Semantic Concept Detection. In Proceedings of IEEE International Conference on Semantic Computing (ICSC) 235-238.

Rajkumar Kannan

Rajkumar Kannan received the B.Sc and M.Sc degrees in Computer Science from Bharathidasan University – Tiruchirappalli, India in 1991 and 1993 respectively and the PhD degree in Computer Science from National Institute of Technology – Tiruchirappalli, India in 2007. Rajkumar works for King Faisal University, Saudi Arabia in the College of Computer Science and Information Technology. His research activities primarily lie at the confluence of multimedia, information retrieval, semantic web, social informatics and collective intelligence. Rajkumar is a member of ACM, CSI-India and ISTE-India.

Gheorghita Ghinea

Gheorghita Ghinea received the B.Sc. and B.Sc. (Hons.) degrees in computer science and mathematics, and the M.Sc. degree in computer science from the University of the Witwatersrand, Johannesburg, South Africa, in 1993, 1994, and 1996, respectively, and the Ph.D. degree in computer science from the University of Reading, Reading, U.K., in 2000. He is a Reader in the School of Information Systems, Computing and Mathematics, Brunel University, Uxbridge, U.K. His current research interests include multimedia computing, telemedicine, quality of service, as well as computer networking and security issues.

Sridhar Swaminathan

Sridhar Swaminathan received his bachelors and masters degrees in Computer Science from Bishop Heber College (Autonomous), Tiruchirappalli, India. He is currently pursuing the Ph.D. degree in Computer Science at the Department of Computer Science, Bishop Heber

College (Autonomous), Tiruchirappalli, India. His main interests are in Computer Vision, Information Retrieval and Machine Learning.