A User-Centric System for Home Movie Summarisation

Saman H. Cooray, Hyowon Lee, and Noel E. O'Connor

CLARITY: Centre for Censor Web Technologies, Dublin City University, Ireland
saman.cooray@dcu.ie

Abstract. In this paper we present a user-centric summarisation system that combines automatic visual-content analysis with user-interface design features as a practical method for home movie summarisation. The proposed summarisation system is designed in such a manner that the video segmentation results generated by the automatic content analysis tools are further subject to refinement through the use of an intuitive user-interface so that the automatically created summaries can be effectively tailored to each individual's personal need. To this end, we study a number of content analysis techniques to facilitate the efficient computation of video summaries, and more specifically emphasise the need for employing an efficient and robust optical flow field computation method for sub-shot segmentation in home movies. Due to the subjectivity of video summarisation and the inherent challenges associated with automatic content analysis, we propose novel user-interface design features as a means to enable the creation of meaningful home movie summaries in a simple manner. The main features of the proposed summarisation system include the ability to automatically create summaries of different visual comprehension, interactively defining the target length of the desired summary, easy and interactive viewing of the content in terms of a storyboard, and manual refinement of the boundaries of the automatically selected video segments in the summary.

Keywords: video summarisation; camera motion estimation; user interaction; home movie

1 Introduction

We are witnessing increasing numbers of video repositories being created by many home users due mainly to the ubiquitous nature of video capture devices that are becoming more affordable. One significant requirement in dealing with the management of such raw video archives is automated video summarisation; a task that can effectively help users organise many hours of movie material, leading to improved viewing experience overall.

Video summarisation, in general, refers to the mechanism of creating a concise version of the original digital video, to facilitate users to perform efficient browsing, search and retrieval of multimedia content in large-scale video archives. In a task like browsing of large video archives, a video summary can provide users with useful information to get a rough idea about the original content of the video in a much shorter time. For summarisation, home movies pose a particularly difficult challenge due to unrestricted capture and lack of storyline present in the content. To create compact movie clips from an unedited video content, home users currently rely on some existing video editing tools, such as Apple's iMovie[1], and Adobe Premiere[2]. Unfortunately, using any of these tools is a laborious and cumbersome process. An automated solution must also ensure that the following principles are satisfied, in order that a summary meets the purpose of real users.

- The most important events and activities are included in the summary.
- Redundant events are sufficiently filtered and excluded from the summary.

Addressing numerous challenges in home movie summarisation, a substantial amount of work has been reported in the literature, and in particular, the last decade has seen greater interest from the research community towards the development of summarisation tools [1–9]. An automatic home video abstraction method was proposed by Lienhart [1] using a date/time based clustering approach and a shot shortening method based on sound features. Kender and Yeo [2] presented a home-video summarisation system based on the use of a zoom-and-hold filter as an implicit human visual attention rule applied when capturing home movies. A probabilistic hierarchical clustering approach was proposed by the authors of [3] using visual and temporal features to discover cluster structure in home video. Huang *et al.* [4] presented an intelligent home video management system using fast-pan elimination, face-shot detection, etc. Recently, Mei *et al.* [5] proposed a novel home video summarisation method based on the exploitation of the user's intention at the time of capture, as a complementary mechanism to existing content analysis schemes. The authors of [6] make use of the home users' photo libraries to infer their preferences for video summarisation. Wang *et al.* [7] presented an information-theoretic approach to content selection as an effective method for selecting the most important content in home video editing. By modeling the co-occurrence statistics between characters (who) and scenes (where), the authors create a compact representation of raw footage from which they extract the most important content using a joint entropy measure. More recently, we presented an interactive and multi-level framework for home video summarisation, combining automatic content analysis with user interaction to create visually comprehensive summaries [8]. Peng *et al.* [9] proposed a user experience model for home video summarisation, taking into account the user's reaction such as eye movement and facial expression when viewing videos. Despite growing interest from the research community, automatic summarisation of home movie remains a challenging research topic due to the presence of unstructured storyline and unrestricted capture in home video footage.

Recently, there has also been a significant body of work on user-interface technologies for home video editing and authoring. The Hitchcock system [10]

[1] urlhttp://www.apple.com/ilife/imovie/

[2] urlhttp://www.adobe.com/products/premiereel/

performs automatic motion analysis of the raw video to determine which parts of the video, i.e. clips, should be included in the summary. Users can interactively override the start and end time of a clip by re-sizing its keyframe. Campanella *et al.* [11] developed the Edit While Watching (EWW) system that has the ability to automatically create an edited version of the raw home video and allow the user to refine results interactively. Upon loading a raw video into the system, a set of short video segments are created based on the analysis of low-level features, such as camera motion, contrast and luminosity, which collectively form the automatically edited version of the video. The system then allows the user to add/remove content to/from the edited version at sub-shot, shot or scene level. Based on the study of existing research and commercial frameworks, it is clear that a practical approach to home movie summarisation should consider a user-centric scenario, ensuring that the work on the part of the user is reduced to the best possible level. Furthermore, user studies carried out by the authors of [12, 13] suggest that home users always want to have the flexibility to tailor the automatically created summaries according to their personal needs.

In this paper, we present a home movie summarisation system, focussing on design concepts of the user-interface while taking into account the challenges associated with automatic content analysis and the subjectivity of video summarisation. The proposed system allows a user to easily create a summarised movie clip composed of the most informative portions of the raw video. The rest of the paper is organised as follows. In Section 2, a description of the proposed home video summarisation system is presented. Section 3 gives a description of sub-shot segmentation, including an experimental analysis of sub-shot segmentation in our framework. Section 4 is devoted to a short description of the summarisation engine. The proposed user-interface design approach is then presented in Section 5. Finally, a conclusion is given in Section 6.

2 Proposed Home Movie Summarisation System

Our home movie summarisation system comprises a number of automatic visual-content analysis techniques as well as a user interaction step as shown in Figure 1. A raw video is fed into the sub-shot segmentation module, which computes the global camera motion parameters of the video and in turn decomposes it into 4 different sub-shot types called pan, tilt, zoom and static, as described in Section 3. Each identified sub-shot is then further processed using the content representation method described in [8]. The resulting sub-shot footprints are input to the summarisation engine, which then performs analysis to identify the most dynamic content as well as the redundant information present in the raw video footage. User interaction functionalities are supported to drive the summarisation process whereby a user will be able to create a particular summary of desired target length whilst being able to interactively view and refine the summarisation results (see Section 5 for details of the user interaction features supported by the system). If the user is happy, she can then confirm and request the final summary be created and saved to disk.

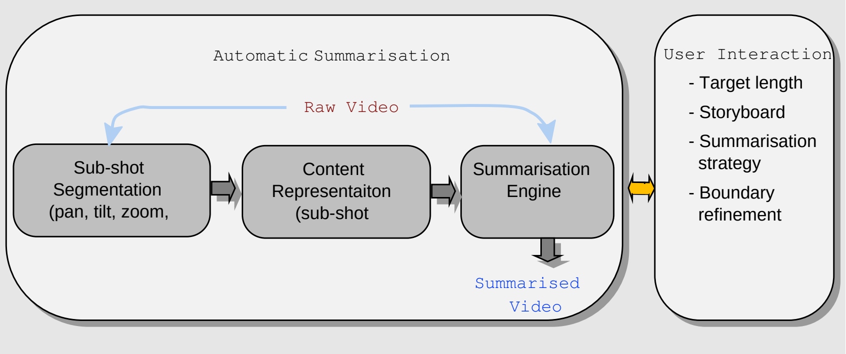


Fig. 1: Proposed home movie summarisation system.

In this paper, we focus primarily on design concepts of the graphical user interface as other aspects of the system are reported elsewhere [8, 14]. Additionally, we present a description of the experimental analysis carried out on sub-shot segmentation, emphasizing the need for identifying an efficient and robust sub-shot segmentation method for home movie summarisation.

3 Sub-shot Segmentation

Employing an effective sub-shot segmentation method is crucial to the success of home video segmentation, which in turn can lead to a significant reduction of subsequent user-interaction required for creating a visually appealing summary for real users. Sub-shot segmentation of home movie footage based on the use of camera motion estimation is, however, a highly computationally expensive task [8, 15]. Thus, identifying an efficient and robust sub-shot segmentation technique is particularly important for the proposed summarisation framework. Although other techniques of much lower computational complexity appear to exist in the literature [16], we believe that detecting sub-shots in line with the change in dominant camera motion enables us to uncover the structure of raw home movie content more effectively.

The flow diagram of the sub-shot segmentation approach employed in our framework is shown in Figure 2. In the pre-processing stage, the raw video (V_r) is first decoded following which each frame of the video is converted to grayscale and resized. Then, the optical flow field is computed for each pair of consecutive frames in the camera motion estimation stage. Fitting those motion vector fields to a 2-D affine model and combining with the RANSAC algorithm, the best transformation between each pair of frames is computed. By comparing the values of each model parameter with a suitably determined threshold, classification of the global motion is carried out for each frame of the video. Finally, a filtering step is applied to all classified frames to determine the type of sub-shots present in the video.

Based on our extensive experiments carried out on home movie summarisation, it was evident that computing the optical flow field corresponds to the

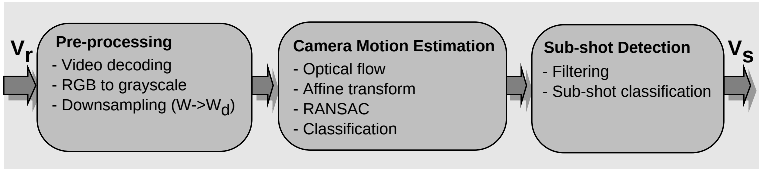


Fig. 2: Flow diagram of sub-shot segmentation: V_r and V_s represent the raw and sub-shot video respectively.

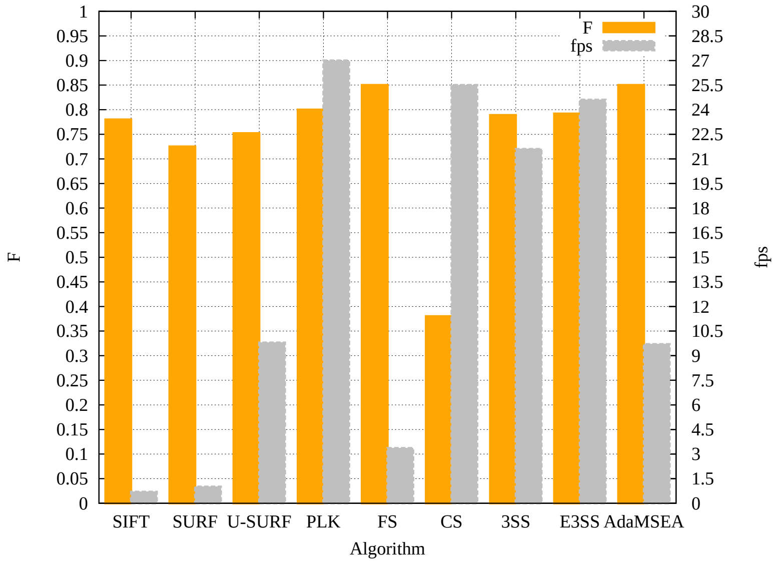


Fig. 3: Performance comparison of sub-shot segmentation based on the use of different optical flow field computation methods.

most time consuming task in sub-shot segmentation. For example, using SIFT features for sub-shot segmentation results in over 90% of the total computational time of summarisation [8]. To this end, we conducted experiments to identify an efficient and robust sub-shot segmentation technique based on the use of different feature-based and block-matching optical flow field computation methods. Feature-based methods include SIFT [17], SURF [18] and Pyramidal Lucas-Kanade (PLK) [19] while block-matching algorithms include full search (FS), cross search (CS) [20], three step search (3SS) [21], efficient 3SS (E3SS) [22] and adaptive multilevel successive elimination algorithm (AdaMSEA) [23]. A test data set consisting of 28 typical home movies (191 sub-shots in total) with varying camera motions, events and durations was used to test the performance of sub-shot segmentation in this experimental study.

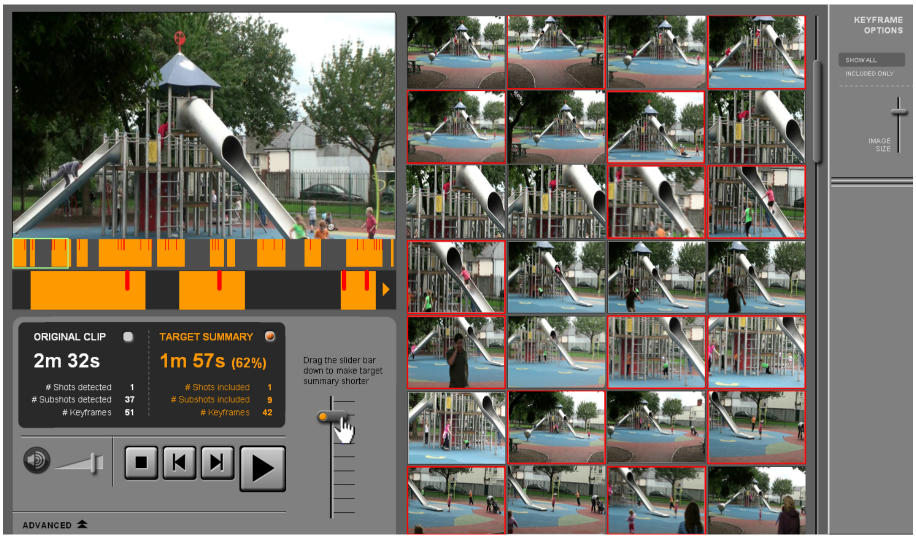
Figure 3 shows a comparison of the different algorithms in terms of accuracy against efficiency using the F-measure and frames-per-second (fps) evaluation criteria. It can be seen that the full-search block matching algorithm and AdaMSEA (which falls into the category of efficient full-search block algorithms) perform best, followed by PLK, E3SS, 3SS, SIFT, U-SURF, SURF and CS in terms of accuracy. There is a 0.05 difference in "F" value between the two best performing algorithms, i.e. full-search (or AdaMSEA) and PLK. However comparing the efficiency figures of AdaMSEA makes it clear that, considering the accuarcy/efficiency compromise home movie summarisation applications require, the PLK algorithm is the preferred optical flow computation choice for sub-shot detection in home movie content. A detailed performance comparison of the algorithms is given in [14].

4 Summarisation Engine

The summarisation engine performs a range of automatic analysis to facilitate the creation of a summarised video. Relying on the content representation method built on the principle of sub-shot footprints [8], the main function of the summarisation engine is to automatically identify which portions of the raw video to be included in the summary, given the user inputs such as target length, summarisation strategy, coverage, etc. It also provides additional information to the user during user interaction stages, such as manually adjusting the boundaries of video segments, changing the target length, and selecting the summarisation strategy.

In order to rank a set of sub-shots based on the significance of content and select the most relevant sub-shots for summarisation, the summarisation engine analyses the coverage and intersection of sub-shot footprints [8]. Based on this analysis, our summarisation framework supports three different summarisation strategies, thereby providing users the flexibility to select the best automatically edited version of a video which can be subsequently refined to tailor for their own needs with minimum interaction. Definitions of the three summarisation strategies provided by the system through the use of an advanced panel (see Figure 5) are as follows. Given target length, T:

- *Strategy 1 (Prominent)*: Arrange sub-shots iteratively so that those corresponding to high coverage and low intersection appear in the top positions of the rank list until their aggregate length ie equal to T.
- *Strategy 2 (Coverage)*: Arrange sub-shots iteratively so that those corresponding to high coverage and low intersection appear in the top positions of the rank list until a pre-defined level of coverage is reached. Then, extract the most dynamic segment from each of those sub-shots so that their aggregate length is equal to T.
- *Strategy 3 (All)*: Extract the most dynamic segment from each of the full set of sub-shots so that their aggregate length is equal to T. This option is set as the default setting, ensuring that every sub-shot is preserved to some degree in the final summary.



ORIGINAL CLIP
2m 32s
Shots detected 1
Subshots detected 37
Keyframes 51

TARGET SUMMARY
1m 57s (62%)
Shots included 1
Subshots included 9
Keyframes 42

Drag the slider bar down to make target summary shorter

Fig. 4: Screen-shot showing initial home movie summarisation results.

5 Interaction Design

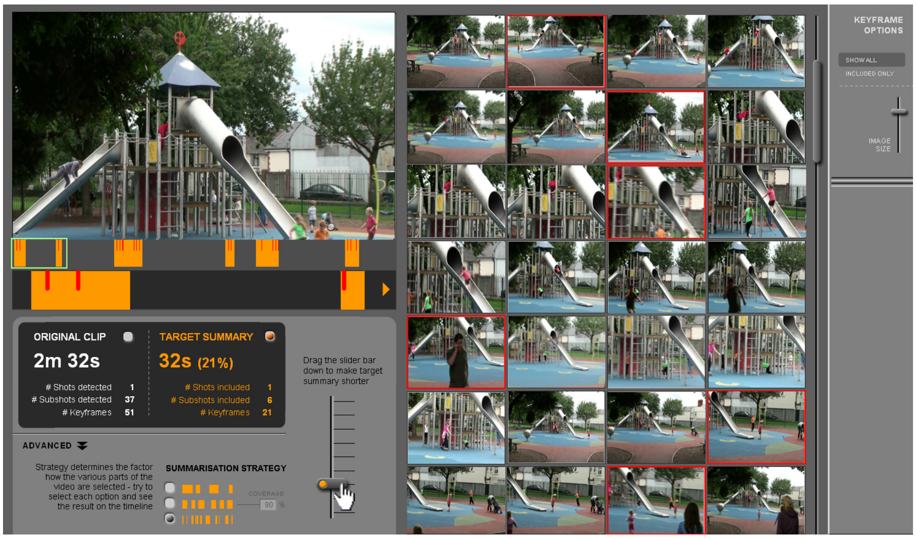
Existing video editing tools, such as iMovie and Adobe Premiere, allow users to perform video editing as a fully manual process. They also inevitably feature complex user-interfaces with a number of layers of timelines to be overlayed, expecting the user manipulates this by adding more segments, adjusting the temporal sequences between the segments, etc. The strength of the proposed video summarisation system is that it simplifies the manual manipulation process by automatically preparing a near optimal summary template, which the user can further refine to tailor for his/her own needs with simple interaction.

The scenario and the front-end user-interaction we have designed is to support a user to easily view the contents, automatically summarise into a compact video clip, and then manually adjust if wished. In this way, automatic summarisation could serve as a pre-edit processing that takes care of most of the otherwise time-consuming, repetitive and labour-intensive editing tasks.

5.1 Initial Summarisation and Browsing Scheme

Figure 4 shows a screen-shot where a home video of a user and her children playing in a playground is loaded into the system, and initially processed and presented to the user on her laptop screen.

On the top-left of the screen is the video playback panel where the user can play the original or summarised video. Below it is a timeline that shows the segments of the video that are included (in orange) and omitted (in dark grey) in the summary. Thin vertical bars marked on the top edge of the timeline (in red) indicate the positions in the original video from where the representative keyframes have been extracted. The timeline is double-layered where the top half


ORIGINAL CLIP ○ 2m 32s
Shots detected 1
Subshots detected 37
Keyframes 51

TARGET SUMMARY ● 32s (21%)
Shots included 1
Subshots included 6
Keyframes 21

Drag the slider bar down to make target summary shorter

Strategy determines the factor how the various parts of the video are selected - try to select each option and see the result on the timeline

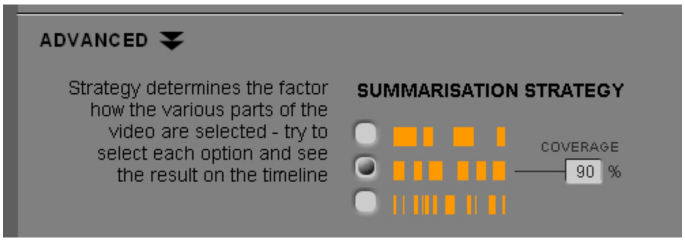
COVERAGE 90 %

Fig. 5: Advanced user control of the summarisation.

shows the full duration of the raw video providing an overview of the summary while the bottom half shows a zoomed-in portion taken from the top half covering the green rectangular area (in this scenario the first 20 seconds of the original video). Presented on the right half of the screen is a "storyboard" of static keyframes. The keyframes marked with red borderlines correspond to the thin red vertical bars on the timeline. The user has an option to view only those selected keyframes (highlighted in red) by selecting the "Included Only" option on the Keyframe Options panel on the right. Also, the size of each keyframe can be changed by dragging up and down the "Image Size" vertical slider bar below.

The current status of summarisation is indicated in the dark text box just below the timeline. In Figure 4, this text box indicates that a "2 minutes and 32 seconds" long raw video has been summarised into a shorter duration of "1 minute and 57 seconds" corresponding to 62% of the original size. At this stage the user can click on the Play button (the larger of the 4 buttons provided below the text box) to view the resultant summary on the player screen that will only play the orange-highlighted parts on the timeline. The other 3 buttons allow the current playback point to jump to the beginning of the next and previous selected segment or stop. Button actions are applied to the summarised video as the yellow radio button beside the label "Target summary" in the text box is selected. When the radio button for the original clip is selected, the buttons that jump to next/previous segment switch to Fast Forward/Backward buttons, in order to allow quicker navigation of the original video clip.

While the above is the initial summarisation outcome automatically generated by the system, the user can now customise the summary in different ways. The vertical slider bar provided just beside the 4 buttons can be dragged up and down in order to change the target length of the summary. For example, as the



ADVANCED

Strategy determines the factor how the various parts of the video are selected - try to select each option and see the result on the timeline

SUMMARISATION STRATEGY

○
●
○

COVERAGE 90 %

Fig. 6: Selecting a summarisation strategy using advanced panel.

user drags the knob down, she will immediately notice the orange-highlighted portions on the timeline becoming thiner and thiner, the number of keyframes with red borders in the storyboard decreasing, and the target duration displayed in the text box becoming smaller. Given a target length, the summarisation engine then performs automatic analysis to determine the most informative portions of the original video to be included in the summary. In this way, our system facilitates users creating video summaries through a simple user intervention process while making best use of the automatic content analysis technologies.

5.2 Advanced Summarisation

A more fine-grained user control of the summarisation process is possible if the user wishes. The "Advanced" button provided at the bottom-left of the screen can be clicked to bring up a small panel where the user can specify different summarisation strategies. Figure 5 shows a screenshot where the Advanced panel has been slided up, and the user dragged down the vertical slider bar in order to reduce the summary duration to only 32 seconds.

A detailed view of the advanced panel is shown in Figure 6, depicting the options available for "summarisation strategy". The 3 options represent (1) include prominent sub-shots, (2) select sub-shots based on coverage, and (3) include as many of all sub-shots. Option 2 is accompanied by an additional setting for the *coverage* where the default value is set to 90%. If the user selects either option 1 or 3, the Coverage setting will be disabled (see Figure 5). Even if the user does not fully understand the technical implications of these settings, she can quickly experiment with these options by observing the immediate changes on the timeline and the storyboard. In general, the user can notice that option 1 tends to generate a summary with a small number of large chunks of the video; a medium number of medium sized chunks for option 2; and for option 3 a large number of small chunks. Any of these options can be used in conjunction with the adjacent vertical slider bar that lets the user set the duration of the summary.

5.3 Summary Customisation: Manual Refinement

If the user is not happy with the above results, she can further customise the summaries through manual refinement. She can simply move the mouse cursor over the timeline and drag the borders between orange (selected portion) and

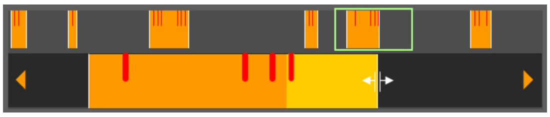


Fig. 7: Manually refining the boundaries of a segment.

dark gray (omitted portion) to manually adjust the boundaries of the video segments. Figure 7 shows an enlarged timeline where the user initially dragged the green rectangular frame on the top half of the timeline to about 3/4 into the video with the bottom half showing a zoomed-in portion of that area. The user is currently extending the orange block by dragging its border to the right, which is indicated by the lighter orange colour area as the portion being added to the summary. As the user adjusts the segment's boundary in this way, she can see the relevant changes being displayed in the text box and the storyboard. In Figure 5, the storyboard highlights 21 keyframes (only 6 of them shown in the top part of the scrollable storyboard) as a result of this adjustment as opposed to the initially highlighted 42 keyframes (and 14 of them shown) in Figure 4.

6 Conclusion

We presented a novel home-movie summarisation framework that combines automatic content analysis with an intuitive user-interface design approach, exploiting the synergy between computer power and human abilities to guarantee the best summarisation results in a simple and efficient manner. We demonstrated the challenges associated with automatic video segmentation through an experimental analysis of sub-shot detection in home movie footage. Using the combined approach, we propose that the issues arising due to the subjectivity of video summarisation and automatic content analysis can be suitably addressed. At present, the proposed user-interface exists at the design stage. By employing an intuitive user-interface design approach that functions in combination with efficient content analysis modules in the back-end, we believe that our approach meets the requirements of real home-movie users. We intend to carry out a number of evaluation tests to assess the effectiveness of our approach based upon the feedback from real users in the future.

Acknowledgment

This research is supported and funded by *BT Group plc*, IRCSET and the *Science Foundation Ireland* under grant 07/CE/I1147.

References

1. Lienhart, R.: Abstracting Home Video Automatically., ACM Multimedia, Orlando, FL, USA, 37–40 (1999).

2. Kender, J. R. and Yeo, B.-L.: On the Structure and Analysis of Home Videos., In Proc. of ACCV, Taipei, Jan. (2000).
3. Gatica-Perez, D. and Loui, A. and Sun, M.-T.: Finding Structure in Home Video by Probabilistic Hierarchical Clustering., IEEE Tran. on Circuits and Systems for Video Tech., 13(6), 539–548 (2003).
4. Huang, S.-H. and Wu, Q.-J.: Intelligent home video management system., Intl. Conf. on Information Technology: Research and Education, 176–180 (2005).
5. Mei, T. and Hua, X.-S. and Zhou, H.-Q. and Li, S.: Modeling and Mining of Users' Capture Intention for Home Videos., IEEE Tran. on Multimedia, 9, 66–77 (2007).
6. Takeuchi, Y. and Sugimoto, M.: User-Adaptive Home Video Summarization using Personal Photo Libraries., In Proc. of CIVR, 472–479 (2007).
7. Wang, P. P. and wang, T. and et al.: Information Theoritic Content Selection for Automated Home Video Editing., In Proc. of ICIP, Texas, USA, 537-540 (2007).
8. S. H. Cooray and H. Bredin and Li-Qun Xu and Noel E. O'Connor, *An Interactive and Multi-level Framework for Summarising User Generated Videos*, ACM MM, pages 685-688, Beijing, China, 2009.
9. Peng, W.-T. and Huang, W.-J. and et. al.: A User Experience Model for Home Video Summarization., In Proc. of MMM, Chongqing, China, 484–495 (2009).
10. Girgensohn, A. and Boreczky, J. and et al.: A Semi-automatic Approach to Home Video Editing., In Proc. of ACM Symp. on User Interface Software and Technology, San Diego, CA, USA, 81–89, Nov. (2000).
11. Campanella, M. and Weda, J. and Barbieri, M.: Edit while watching: home video editing made easy., In Proc. of SPIE, vol. 6506, 65060L (2007).
12. Wu, P. and Obrador, P.: Personal Video Manager: Managing and Mining Home Video Collections., In Proc. of SPIE, Bellingham, vol. 5960, 775–785 (2005).
13. Salton, G and Singhal, A. and et al.: Automatic Text Struturing and Summarization., Information Processing and Management, 22(2), 193–207 (1997).
14. S. H. Cooray and N. E. O'Connor, *Identifying an Efficient and Robust Sub-shot Segmentation Method for Home Movie Summarisation*, Submitted to 10^{th} IEEE Intl. Conf. on Intelligent Systems Design and Applications, Nov. 29 - Dec. 1, 2010.
15. L.-X. Tang and T. Meo and X.-S. Hua, *Near-Lossless Video Summarisation*, ACM MM, pages 1049-1052, Beijing, China, 2009.
16. L. Bai and S. Lao and et al., *Automatic Summarization of Rushes Video Using Bipartite Graphs*, In Proc. of SAMT, Koblenz, Germany, Pages: 3 - 14, 2008.
17. D. G. Lowe, *Distinctive image features from scale-invariant keypoints*, Intl. Journal of Computer Vision, pages 91-110, 2004.
18. H. Bay and A. Ess and et al., *SURF: Speeded Up Robust Features*, Computer Vision and Image Understanding (CVIU), vol. 99, no. 3, pages 346-359, 2008.
19. J.-Y. Bouguet, *Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm*, Part of OpenCV library.
20. M. Ghanbari, *The Cross-Search Algorithm for Motion Estimation*, IEEE Tran. on Communications, vol. 38, no. 7, pages 950-953, 1990.
21. T. Koga and K. Linuma, *Motion Compensated Interframe Coding for Video Conferencing*, In Proc. Nat. Telecomuunication Conf., pages G5.3.1–G5.3.5, 1981.
22. X. Jing and L.-P. Chau, *An Efficient Three-Step Search Algorithm for Block Motion Estimation*, IEEE Tran. on Multilmedia, vol. 6, no. 3, pages 435-438, 2004.
23. S.-W. Liu and S.-D. Wei and S.-H. Lai, *Fast Optimal Motion Estimation Based on Gradient-Based Adaptive Multilevel Successive Elimination*, IEEE Tran. on Circuits and Systems for Video Technology, vol. 18, no. 2, pages 263-267, 2008.