Asian Journal of Information Technology 15 (16): 2763-2769, 2016
ISSN: 1682-3915

Proposing of Intelligent Movie Editor

IsraaHadi Ali and Roa'a M. Al_Airaji
Information Technology College, Babylon University, Babylon, Iraq

Abstract: Digital video has become available everywhere. Yet, there is difficulty in the processof video editing because the video is time based medium, dual tracks medium and the available tools make users work at low level of details. This study proposed an intelligent editorto edit video content into short clips and then order the clips on a timeline to create a finished video. The main aims of this study are to simplifying process of video editing and resolve the problem of video and audio overlay. The proposed editor applies the following steps: First, segmentation of video objects. Second, detection and tracking of specific objects based on specificfeatures. Third, put the specific objects in new frames. Fourth, apply functions of the editor such as enhancement of the specific objects, rectification of objectstrajectory, adding special effects etc. Fifth, create new video for the specific objects by merge the new frames that contain it.

Key words: Digital, video editing, video object, segmentation, specific objects

INTRODUCTION

Digital video is becoming available everywhere. Introduced severalof digital video cameras to our daily life and computers are becoming more ability to turn video editing software. There is large number of videos are available in the digital libraries and on the internet. Many of researchers are investigating in their research projects how to seek, abridgement and visualize video but there is few action on modern methods to support the use of the video in addition to playing it. Additional to edit images of the video file, other researchers worked on editing of image that found inside the pdf file. Tawfiq A. Al-Asadiand.

One of the most important point when one produce editor is to protect the copyright of digital product, this is by using watermarking. Abbas and Jawad (2013). Video editing is difficult than text editing. Create a report using video exist in a digital libraries or a new composition using new footage video need to much time and effort than creating an identical report using quoted text or composition by authoring new text (Myers *et al*., 2001).

There are several challenges for video editing are: First, video is a time-based. Which makes browsing and skimming video are difficult. Often, to find the desired clip requires a linear search in source video. Second, digital video is a dual track of video and audio. Synchronize between these tracks is required, also the user should be able to overlay of the video and audio during transitions from one footage toanother. Further, audio and video overlay should disentangle when cutting footage from video for use it elsewhere. Third, the users need to edit video footages and audio words or audio sentences, the available tools force the users to manipulate video at low level of details and audio using waveform to achieve most editing functions such as footage cut, insert text,The user should be accurately identifies particular frames which may include zooming and other several operations. Manipulating aparticular word or sentence using a waveform is also tedious (Casares *et al*., 2002).

Literature review

Related work: Juan Casares and others in 2002 presented "Simplifying Video Editing Using Metadata" the researchers presented editor known Silver. "The silver provides several views with multiple semantic contents" involve timeline views and editable transcript. Silver provides intelligent editing functions that aid the users to manipulate the inconsistence that appear because of the different boundaries in audio and video (Casares *et al*., 2002). Chris Long and others, in 2003 presented "Video Editing Using Lenses and Semantic Zooming". The researchers proposed system provides several lenses on the same timeline, this make user can see several positions at the same time.

Hongcheng Wang and His colleagues in 2004 presented "Seamless Video Editing". The researcher developed a new approach to editing of video in the gradient domain, they create a new gradient field by altering or mixing of the spatio-temporal gradient fields of

Corresponding Author: IsraaHadi Ali, Information Technology College, Babylon University, Babylon, Iraq

specific video, they proposed system to solve problems of spatial consistency and temporal coherency. Zigelbaumand *et al.* (2007) presented "The Tangible Video Editor: Collaborative Video Editing with Active Tokens".

The researchers used in their approach several handled computers to encourage the users to collaborative work. They used active tokens to provide flexible interface and make users enable to organize contents of interface. They used number of methods to improve projection-based tabletop interfaces.

Video editing: There are three sub-definitions of video editing. Video editing can refer to linear video editing, non-linear video editing and vision mixing.In general, video editing can refer to processing and altering video shots to create newwork. Processing and altering involve cutting segments, adding audio clips, applying re-sequencing videoclips,creatingtransitionsbetween clips, enhancements and inserting special effects (Ache *et al*., 2008).

Linear video editing: Includes the essential operation of choosing clips, reordering clips and altering the audio and images on clips all by using a video tape. This is done by Appling cut and splice on video tape to create a new series.This process very difficult and tedious (Peng *et al*., 2012).

Non-linear video editing: The fundamental notion for the process of video editing remains the same in non-linear video editing but with different tools. The computer-based cutting and pasting of clips eliminate the destructive nature of linear video editing. The original recording of the film remain without tainting or altering and copy of it used in editing. Because the original recording never touches, it is remains intact even with multiple alterations. At any time, the editor can act on any portion of the film. This is a way faster method and the video editing software inexpensive (Acha *et al*., 2008).

Vision mixing: Production switching is the second name for vision mixing. This type typically used in TV telecast. The process includes choosing video from multiple sources and featuring it on the telecast. Occasionally, numerous of video sources combined with each other and some special effects inserted to it to achieve a dynamic vision for a telecast (Acha *et al*., 2008).

The most fundamental aim of video editing is eliminate of undesirable video clips. After eliminating the faults of the video, one can attempt and create new film that contains only clips are what you want to be in the new creation. When this aim achieved, the next aim is inserting special effects, impressive transitions, beautiful music and other wonderful imagery. The last aim is giving a meaningful sense for the new work.

Video object segmentation: Video object segmentation is a necessary process in many video applications. It is required for video editing and special effects whenever objects must deleted, moved, individually editedor layered. It also used in object recognition, compression and 3D reconstruction from video (Lee, 1981; Lee *et al*., 2011). There are two manners for video object segmentation are interactive and supervised. In interactive methods, boundaries of object are annotate in some primary frames, then deployed to other frames while a user stands by to correct mistakes (Yang *et al*., 2010). Tracking-based methods try to minimize the supervision by applying segmentation on the first frame only. All these methods require user input the regions that represent object and may sensitive to a user's annotation expertise (Price *et al*., 2009).

Bottom-up methods can apply video segmentation in a fullyautomatic way, based on cues like appearance similarity and motion. Motion segmentation approachesusing bottom-up motion cues to cluster points in video frames (Shi and Malik, 1998). Recent approaches applying segmentation either in pixel-level of spatio-temporal video volume from scratch, start by applying segmentation on each frame and then match segments across nearer frames (Vazquez-Reina *et al*., 2010) or cluster long-term objects trajectory by using dense flow without any top-down notion of objects, however such methods tend to over-segment, yielding regions that taken alone may lack semantic meaning (Brox and Malik, 2010).

The goal of image segmentation algorithms is to division the image into meaningful and homogeneous areas. Each segmentation algorithm handles two issues, the approach for performing efficient division and the criteria for a good division. Figure 1 explains segmentation techniques that are relevant to video object:

Mean-shift clustering: In mean shift clustering, clusters found in the spatial location and color space. Given an image, the process started by randomly selecting a large set of clusters centers from the data (Fig. 2). Then, shift every cluster center to the mean of the data. The mean of the data is lying inside the multi-dimensional ellipsoid centered on the cluster center. Mean-shift vector is a

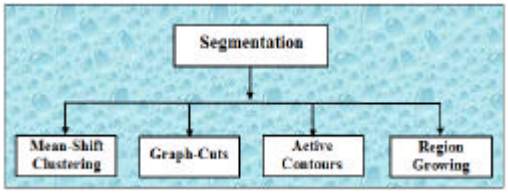

Fig. 1: Segmentation techniques

Fig. 2: Segmentation of the image: a) Original image; b) Using mean-shift segmentation

vector which defined by the old and the new cluster centers. The mean-shift vector calculated repetitively until there is no change in the locations of cluster centers. Note that some clusters may bemerging during the mean-shift repetitions (Comaniciu and Meer, 2002). Figure 2b explains the segmentation using the mean-shift method.

Mean-shift clustering is applicable in image regularization, edge detection and tracking (Comaniciu and Meer, 2002; Comaniciu *et al*., 2003). To get the best segmentation, Mean-shift based segmentation need to fine tuning of several parameters, for example, chosen of the color, chosen of spatial kernel bandwidths and the threshold that minimize the size of the area.

Image segmentation using graph-cuts: In graph cut method the segmentation of the image considered as the graph-partitioning problem where the G is the graph (image) and V = {u, v…} are vertices (pixels) of a graph (image). This vertices (pixels) are divided into N sub-graphs (regions) where Ai, Ai n Aj = Ø, i j, by pruning the weighted edges of the graph. The weight of edges iscalculating by brightness, texture or color similarity between the nodes. For two subgraphs, the cut is the total weight of the pruned edges. The aim of Wu and Leahy is to find the divisions that minimize a cut. So they used the minimum cut criterion. In their approach, they used color similarity to calculate the weights. Limitation of minimum cut method is its bias toward over segmenting the image. This is because of the increase in cost of a cut with the number of edges going across the two partitioned segments (Wu and Leahy, 1993; Shi and Malik, 1998) proposed the normalized cut to resolve the segmentation problem. In their approach the cut based on

Fig. 3: Segmentation of the image: a) Original image; b) Using normalized cuts segmentation

both, the total weights of edge in the cut and the ratio of the sum connection weights of node in every division to all nodes of the graph. For image-based segmentation, the weights are calculating by multiplication of the spatial proximity and the color similarity between the nodes. When the weights are calculated between each pair of nodes, a weight matrix W and a diagonal matrix D: Where $D_{i,i} = \sum^{N} W_{i,j}$. to perform the segmentation, the eigenvectors and the eigenvalues of the generalized eigensystem (D-W)y = λ Dy must be calculated , then partition the image into two segments by using the smallest eigenvector . This process is iteratively for every new segment until reach to specific threshold. Fig. 3b, show the results of segmentation using the normalized cuts method.

In normalized cuts-based segmentation, the solution to the generalized eigensystem for large images can be expensive in terms of processing and memory requirements. However, this method requires fewer manually selected parameters, compared to mean shift segmentation. Normalized cuts also used in the context of tracking object contours (Xu and Ahuja, 2002).

Region growing approach: Is an approach start with a number of seed pixels. These pixels sign every of the objects to be segmented. The areas are expanding from these pixels by allocated those neighboring pixels that have similar features to every seed pixel. The similarity is measure by different features like color, texture, intensity, gray level, etc., the pixel with highest similarity measured this method is allocated to the respective area. This process stopswhen all pixels allocated to an area. Because seeded area expanding demands seeds as extra input, the results of this approach are rely on the select of seeds and the seeds may poorly place because of the presence of noise in the image (Kumar and Srinivas, 2013).

Active contours: To achieving object segmentation using active contour method, a closed contour evolved to the object's boundary, so that the contour tightly encloses the object area. The energy function defines the fitness of

the contour which governs the evolution of the contour to the precise object area (Caselles *et al*., 1995). The following common form shows the energy function for contour evolution:

$$E(T)=\int_0^1 E_{int}(V)+E_{im}(V)+E_{ext}(V)ds$$

Where s refer to the arc-length of the contour, E_{int} involves regularization constraints which are curvature term, first-order (Δv) continuity term or second-order ($\Delta^2 v$) continuity term to find the shortest contour, E_{im} involves appearance-based energy, Image-based energy, E_{im}, can be calculated locally or globally. Local information is evaluatearound the contour and is in the form of an image gradient. In contrast, global features calculated inside and outside of the object area. Global features involve Color and texture and E_{ext} involves additional constraints (Paragios and Deriche, 2002; Yilmaz and Shan, 2004).

MATERIALS AND METHODS

Figure 4 explain work steps of the proposed method The proposed method consists of many steps as follows:

First step: In this step we open AVI structure and get list movi (still images/sequence frames), then get information such as total of frame width of the frame and height of the frame from AVI header then split the movie film into frames (still image).

Second step: Apply segmentation algorithm on the first frame only. The aim of segmentation isto detect the regions that represent objects and forclassifying all pixels in a given frame as an object or non-object pixels. In this study region-growing segmentation is used to detect the object regions.the region grown method start with number of seed points , then the regions from these seed points are growing by allocate those neighboring pixels that have similar features with the seed point to the respective region, the features that used to measure the similarity are texture, color, intensity and gray level.

After that determine target object based on features such as texture, size, boundary, etc and extract features for target object such as edge, area, center and interest points, these features are useful and require tousing in the tracking target object, the target object is putting in new frame, this step is required to construct new movie.

Then after that, apply tracking techniques on the other frames to track the target object through the movie.

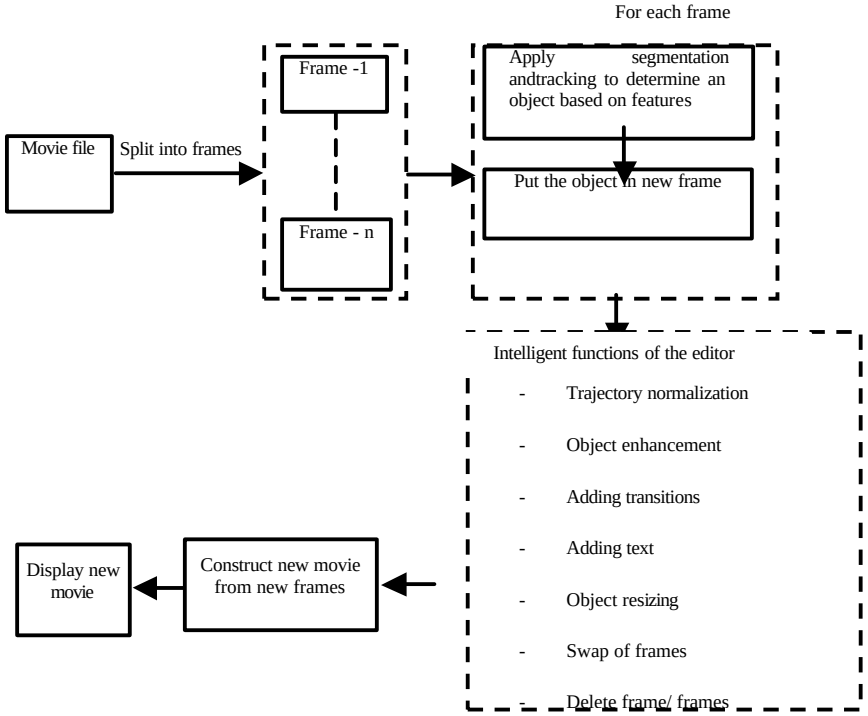

Fig. 4: Block diagram of the proposed method

a b

Fig. 5: a) Original image; b) Result of inpaiting

Center of the target object that have been obtained in the first frame is used to track the object in the second frame, then new center is calculated for the target object in the second frame to use it to track the target object in the third frame and so on to the last frame. The tracking process starts from the second frame, find region that represent the object through the center which wasobtain from the previous frame and collect the neighbor pixels related to the object, then calculating the new center for this object that use to find the object in the next frame, then putting the object in new frame.

To compute the center point of the object must first compute the area of an object, equation bellow explain calculate area of an object. Where O (x, y) represents the pixel of object. The center of object calculate by the following equation:

$$Xc = \left(\sum_{x} \sum_{y} \left(xo(x,y) \right) \right) / \text{Area}$$

$$Y_c = \left(\sum_{x} \sum_{y} \left(yo(x,y) \right) \right) / \text{Area}$$

Where X_c and $Y_c$are the centroid coordinates of an object.

Third step: In this step intelligent functions may be applied on new frames that contains target object or on the target object itself, these functions enhancement and give new movie beautiful look. The intelligent functions are.

Inpainting: Is the process of altering the image values of the pixels to repair damaged parts of an image such as scratches in an old photograph or to delete unwanted elements in the image. Example bellow explains inpainting process (Fig. 5a,b).

Trajectory normalization: Is the process of rectification of object trajectory by organizing and altering the points of the trajectory to reduce point's redundancy.

Object enhancement: This function plays vital role in improving the appearance of an object by adjusting its features such as brightness, contrast, hue, saturation and speed

Adding transitions: Adding transitions between shots to create flow between them, there are many types of transitions are cut, mix, dissolve, fade, cross-fade, wipe and digital effects:

- Adding text: adding text to the movie to create title cards, captions to the object and anything
- Other functions such as object resizing, swap between frames to rearrange them, delete frame frames, display specific object in multi window change of background cut any object in any frame and paste in another frame, merge number of frames in any movie with number of frames in another movie for contrast new movie and other special effects given beautiful look for new movie

Fourth step: The construct new film to specific object by merging frames that contains it.

RESULTS AND DISCUSSION

Numbers of video file movies was taken are implemented in the proposed method , these movies different in the total of frames, dimensions of image and number of objects, these movie implemented with proposed method and some functions are applied on this movie give the results which explain bellow. In example one an illustration in fig. 6 the new movie in created with new background.

In example two an illustration in Fig. 7 some functions applied on the movie after segmentation and tracking, these functions are object resize to minimizing the size of object, then after create new film for the object, this object is display in four window .

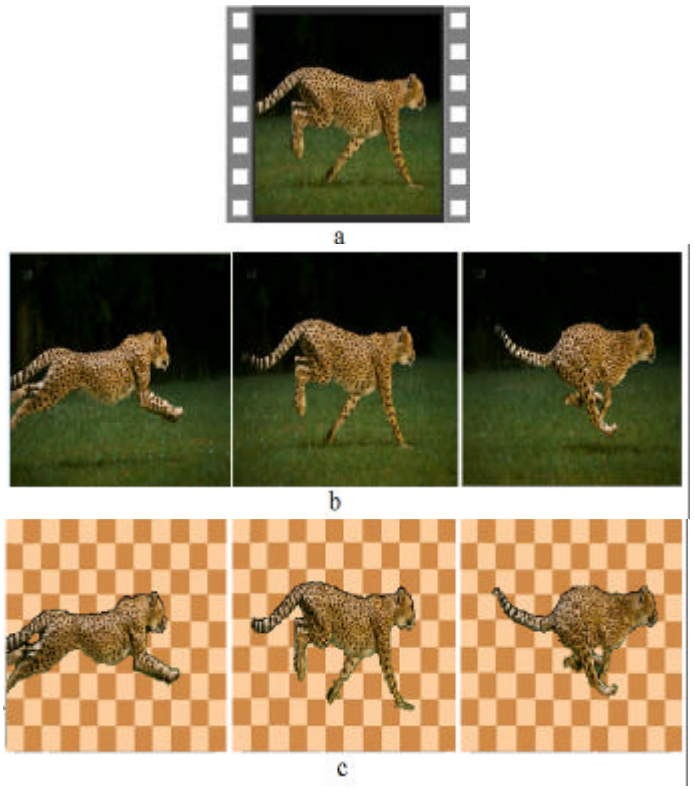

Fig. 6: (a Origin movie; b) Samples of frames for the origin movie; (c Samples of frames for the new movie

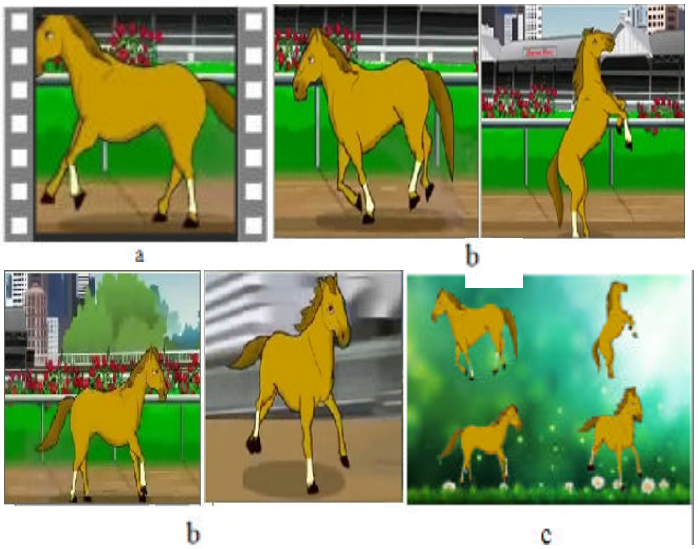

Fig. 7: a) Origin movie; b) Samples of frames for the origin movie; c) Samples of frames for the new movie display in four windows

CONCLUSION

Intelligent movie editor presented in this study make the transition from traditional editor to intelligent editor. Where it is characterized by intelligent behavior and high precision in implementation of its functions,Intelligent editor make editing of video is easier as editing of text.

REFERENCES

Abbas, T.A. and M.J. Jawad, 2013. Proposed an intelligent watermarking in gis environment. J. Earth Sci. Res., 1: 1-5.

Acha, A.R., P. Kohli, C. Rother and A. Fitzgibbon, 2008. Unwrap mosaics: A new representation for video editing. Proceedings of the Conference on ACM Transactions on Graphics (TOG), August 11-15, 2008, ACM, New York, USA., ISBN:978-1-4503-0112-1, pp: 779-786.

Brox, T. and J. Malik, 2010. Object Segmentation by Long Term Analysis of Point Trajectories. In: European Conference on Computer Vision, Daniilidis, K., P. Maragos and N. Paragios (Eds.). Springer, Berlin, Germany, ISBN: 978-3-642-15555-0, pp: 282-295.

Casares, J., A.C. Long, B.A. Myers, R. Bhatnagar and S.M. Stevens *et al*., 2002. Simplifying video editing using metadata. Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, June 25-28, 2002, ACM, New York, USA., ISBN:1-58113-515-7, pp: 157-166.

Caselles, V., R. Kimmel and G. Sapiro, 1995. Geodesic active contours. Proceedings of the 5th International Conference on Computer Vision, June 20-23, 1995, Cambridge, MA., pp: 694-699.

Comaniciu, D. and P. Meer, 2002. Mean shift: A robust approach toward feature space analysis. IEEE Trans. Patt. Anal. Machine Intel., 24: 603-619.

Comaniciu, D., V. Ramesh and P. Meer, 2003. Kernel-based object tracking. IEEE Trans. Pattern Anal. Machine Intell., 25: 564-577.

Kumar, S. and R. Srinivas, 2013. A study on image segmentation and its methods. Intl. J. Adv. Res. Comput. Sci. Software Eng., 3: 1112-1114.

Lee, J.S., 1981. Refined filtering of image noise using local statistics. Comput. Graph. Image Proces., 15: 380-389.

Lee, Y.J., J. Kim and K. Grauman, 2011. Key-segments for video object segmentation. Proceedings of the IEEE International Conference on Computer Vision, November 6-13, 2011, Barcelona, Spain, pp: 1995-2002.

Myers, B.A., J.P. Casares, S. Stevens, L. Dabbish and D. Yocum *et al*., 2001. A multi-view intelligent editor for digital video libraries. Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries, June 24-28, 2001, ACM, New York, USA., ISBN:1-58113-345-6, pp: 106-115.

Paragios, N. and R. Deriche, 2002. Geodesic active regions and level set methods for supervised texture segmentation. Int. J. Computer Vision, 46: 223-247.

Peng, Y., A. Ganesh, J. Wright, W. Xu and Y. Ma, 2012. RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images. IEEE. Transac. Pattern Anal. Mach. Intell., 34: 2233-2246.

Price, B.L., B.S. Morse and S. Cohen, 2009. Livecut: Learning-based interactive video segmentation by evaluation of multiple propagated cues. Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, September 29-October 2, 2009, IEEE, New York, USA., ISBN:978-1-4244-4420-5, pp: 779-786.

Shi, J. and J. Malik, 1998. Motion segmentation and tracking using normalized cuts. Proceedings of the 6th International Conference on Computer Vision, January 7-7, 1998, IEEE, New York, USA., ISBN:81-7319-221-9, pp: 1154-1160.

Vazquez-Reina, A., S. Avidan, H. Pfister and E. Miller, 2010. Multiple hypothesis video segmentation from superpixel flows. Proceedings of the 11th European Conference on Computer Vision, September 5-11, 2010, Heraklion, Crete, Greece, pp: 268-281.

Wu, Z. and R. Leahy, 1993. An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. IEEE. Trans. Pattern Anal. Mach. Intell., 15: 1101-1113.

Xu, N. and N. Ahuja, 2002. Object contour tracking using graph cuts based active contours. Proceedings of the International Conference on Image Processing, September 22-25, 2002, IEEE, New York, USA., ISBN:0-7803-7622-6, pp: 277-280.

Yang, J., B. Price, X. Shen, Z. Lin and J. Yuan, 2016. Fast appearance modeling for automatic primary video object segmentation. IEEE. Transac. Image Proc., 25: 503-515.

Yilmaz, A., X. Li and M. Shah, 2004. Contour-based object tracking with occlusion handling in video acquired using mobile cameras. IEEE. Transac. Pattern Anal. Mach. Intell., 26: 1531-1536.

Zigelbaum, J., M.S. Horn, O. Shaer and R.J. Jacob, 2007. The tangible video editor: Collaborative video editing with active tokens. Proceedings of the 1st International Conference on Tangible and Embedded Interaction, February 15-17, 2007, ACM, New York, USA., ISBN:978-1-59593-619-6, pp: 43-46.