This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LSP.2016.2573042, IEEE Signal Processing Letters

An Efficient Framework for Automatic Highlights Generation from Sports Videos

Ali Javed, Khalid Bashir Bajwa, Hafiz Malik, Senior Member IEEE, and Aun Irtaza

***Abstract*— This paper presents a framework for replay detection in sports videos to generate highlights. For replay detection, the proposed work exploits the following facts: (i) broadcasters introduce gradual transition (GT) effect both at the start and at the end of a replay segment (RS), and (ii) absence of score-captions in a replay segment. The dual-threshold-based method is used to detect GT frames from the input video. A pair of successive gradual transition frames is used to extract the candidate replay-segments. All frames in the selected segment are processed to detect score-caption (SC). To this end, temporal running average is used to filter out temporal variations. First- and second-order statistics are used to binarize the running average image, which is fed to OCR stage for character recognition. The absence/presence of SC is used for replay/live frame labeling. The SC detection stage complements the GT detection process, therefore, a combination of both is expected to result in superior computational complexity and detection accuracy. The performance of the proposed system is evaluated on 22 videos of four different sports (e.g. Cricket, tennis, baseball, and basketball). Experimental results indicate that the proposed method can achieve average detection accuracy ≥ 94.7%.**

***Index Terms*— Gradual transition, highlights, replay detection, score-caption, temporal running average.**

I. Introduction

THE increasing amount of multimedia content available in the cyberspace have sparked research activities to develop efficient video analysis and content management techniques. Analysis and consumption of available videos in the cyberspace is a challenging task for both computing machines and humans. Video summarization techniques are commonly used to address this issue by providing abstract video of the full length videos. There is a growing need for effective video summarization techniques that can provide all the significant events to the consumers in a succinct manner. Video summarization approaches have applications in various domains including sports [1], surveillance [2], healthcare [3], home videos [4], news [5], entertainment [6], etc.

Everyday sports broadcasters generate a massive collection of video content consisting of majority of redundant events and a very few significant events. Video summarization is used to extract significant (or key) events from a full length video. Existing sports video summarization approaches can be divided into (i) summarization from live videos [7]-[9], and (ii) summarization using replay detection [10]-[12]. Ekin *et al.* [1] and Dian *et al.* [13] have combined both live- and replay-based summarization approaches. During live sports broadcasting, replays are commonly used to emphasize on the occurrence of significant events, which is motivation behind using replays for highlights generation from sports videos. Replays, in general, are included after any interesting event in the game to present details of key events in slow motion. It is, therefore, commonly used in sports video analysis for event detection and highlights generation [11]-[18].

Existing replay detection approaches can be classified into (i) learning-based approaches [10]-[12], [14], and (ii) non-learning based approaches [15]-[17]. For example, Pan *et al.* [11] proposed a learning-based framework for logo detection in scene transitions. The method [11] first detects two replay segments (RS) that are used to detect a pair of similar frames in the preceding frames of the detected RS by grouping logo frames. Accuracy and reliability of candidate RS detection is one of the limitations of this method. Such techniques, e.g., logo-detection-based approaches rely on extensive training of the classifier for various logos. In addition, performance of such techniques also depend on the accuracy of logo detection which is a challenging task given variations in logo design, shape, color, size, and placement among different sports, tournaments, and broadcasters. Existing techniques also rely on replay structure [11] and motion features [19]-[20]. For example, Duan *et al.* [20] have used the features of motion variations in support vector machine classifier to detect replays in sports videos.

To address limitations of learning-based methods such as computational complexity, non-learning-based techniques [15]-[16] have been proposed. For example, Nguyen *et al.* [16] used histogram difference and contrast features, and Xu *et al.* [21] computed the accumulative difference of frames to identify the logo frames for replay detection. Performance of these methods depend on the presence of logo frames.

Ali Javed, Khalid Bashir Bajwa, and Syed Aun Irtaza are with Software Engineering Department, FT&IE, UET Taxila, Pakistan.(e-mail: ali.javed@uettaxila.edu.pk). Hafiz Malik is with ECE Department, UM-Dearborn, MI, USA. (e-mail: hafiz@umich.edu).

Similarly, Eldib *et al.* [22] and Chen *et al.* [23] have used statistical features to detect the replay sequences.

To address limitations of existing replay detection methods such as computational complexity of logo detection, camera variations, replay speed, logo design, size, placement, etc., a computationally efficient hybrid method is proposed for automatic highlights generation from sports videos. The main contribution of this paper is to develop a computationally efficient hybrid technique to detect replays for video summarization. It has been observed that broadcasters omit the score-captions (SC) in RSs. Moreover, replay frames contain multiple gradual transitions (GTs). The present work exploits these two observations for replay detection. More specifically, the proposed method uses GTs and SCs for replay detection that is then used for highlight generation. To achieve this goal, dual-threshold-based method [24] is used for GT detection. Detected GT frames are used to extract candidate RSs. Candidate RSs are used for SC detection. The estimated SC is used to discriminate between replay and live video frames. The proposed system is robust to camera variations, replay speed, logo design, size, placement, etc., score captions type, sports broadcasters, and sports category. The performance of the proposed system is evaluated on a dataset of four different sport categories. Experimental results indicate that the proposed system achieves the detection accuracy >= 94.7% averaged over all videos.

II. PROPOSED SYSTEM

The proposed system is divided into two main stages, GT detection and SC detection. The block diagram of the proposed system is shown in Fig. 1.

A. *Gradual Transition (GT) Detection*

Replay segments in sports videos include various types of GTs such as dissolves, wipes, fade-in/out etc. It has been observed that replays in sports videos are sandwiched between GT frames and do not contains SCs. The characteristics of multiple GTs are therefore used to identify the boundaries of a RS by detecting logo frames.

Thresholding of histogram difference between frames of luminance component (i.e., grayscale representation) is used to detect GT. To this end, a dual-threshold is used for thresholding of successive and accumulative histogram differences of luminance component. Here, start of GT is detected by comparing histogram difference of successive frames against a computed threshold T_L [24], and the end of GT is detected by comparing accumulative histogram difference against a computed threshold T_U [24]. More specifically, if successive histogram frame difference lies below T_L and the accumulative histogram frame difference exceeds T_U, then this segment is selected as a possible candidate for GT. If separation between start and end of GT frame-indices is ≥ N_{GT} then a candidate segment is labeled as a GT.

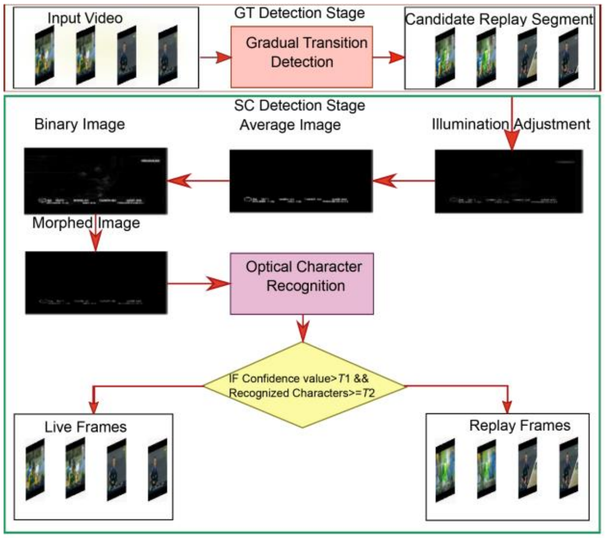

Fig. 1. Block diagram of the proposed system.

1) Candidate Replay Segment (RS) Detection:

Separation between two successive GTs (in number of frames) is used to generate a candidate replay segment. Let S_i and E_i denote start and end of i^{th} GTs, and N_R represents separation between frame indices of S_i and $E_{(i+1)}$. Then a segment between two successive GTs is labeled as a candidate replay segment if it satisfies the following condition, i.e.,

$$2N_{GT} + N_{RL} \le E_{(i+1)} - S_i \le 2N_{GT} + N_{RU} \quad (1)$$

Where N_{RL} and N_{RU} represent lower and upper limits of a replay duration (in number of frames).

To test effectiveness of this approach, we applied it on the selected video dataset. Shown in Fig. 2 is the start and end of candidate replay segments for three videos.

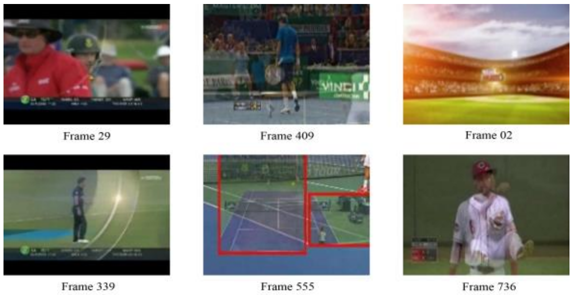

Fig. 2. Top: Start transition frames, Bottom: End transition frames.

B. *Score-Caption Detection*

The SCs are displayed at fixed locations in almost all sports videos. It has been observed through watching extensive amount of sports videos that replay segments do not contain SC. Therefore, SCs are used for replay detection. To this end, only candidate RSs are analyzed to extract SCs. The presence/absence of SC is used to detect replay and live frames.

1) Preprocessing: The preprocessing stage transforms the candidate RSs into a sequence of grayscale images. To reduce computational cost, sequence (of grayscale images) is down-sampled by a factor of 2. Each image is processed for illumination adjustment using the top hat filtering [25]. The top hat filter performs morphological opening with a structuring element *SE* followed by subtraction from the original image. These operations can be expressed as follows:

$$I_{thin}^{(i)} = I^{(i)} \otimes SE \quad (2)$$

$$I_{adj}^{(i)} = I^{(i)} - I_{thin}^{(i)} \quad (3)$$

Where $\mathbf{I_{thin}^{(i)}}$, $\mathbf{I_{adj}^{(i)}}$, and $\mathbf{I^{(i)}}$ represents the thinned image, illumination adjusted image, and input grayscale image respectively, if i^{th} frame. *SE* is the *disk* shaped structuring element of size α, and $\otimes$ is thinning operator.

2) Temporal Running Averaging: A sliding window of length *L* frames is used to compute temporal running average sequence and can be expressed as:

$$I_{avg}^{(i)} = (I_{avg}^{(i-1)} - I^{(i-1)} + I^{(i+1)}) \,/\, L \quad (4)$$

Where $\mathbf{I_{avg}^{(i)}}$ represents average if i^{th} frame.

3) Image Binarization: First- and second-order statistics are computed for average image, $\mathbf{I_{avg}^{(i)}}$, that are used to convert $\mathbf{I_{avg}^{(i)}}$ into binary image using eq. (5):

$$I_{bin}^{(i)}(x,y) = \begin{cases} 0\,, if\ (\mu_i - p * \sigma_i) \le I_{avg}^{(i)}(x,y) \le (\mu_i + p * \sigma_i) \\ 1, \qquad\qquad\qquad\qquad\qquad\qquad otherwise \end{cases} \quad (5)$$

Where $\boldsymbol{\mu_i}$ and $\boldsymbol{\sigma_i}$ represent the mean and standard deviation for $\mathbf{I_{avg}^{(i)}}$, and *p* is a positive real constant.

4) Morphological Thinning: To get rid of outliers, a single pass of morphological thinning is applied on the resulting binary image that can be expressed as:

$$I_{thin}^{(i)} = I_{bin}^{(i)} \otimes SE \quad (6)$$

where $\mathbf{I_{thin}^{(i)}}$ represents thinned image if i^{th} frame.

5) SC detection using Optical Character Recognition (OCR): To recognize contents of SC, the OCR process is applied on the thinned image. The OCR algorithm recognize characters with a confidence. The confidence score associated to each character along with number of characters recognized are used for SC detection. More specifically, (*if confidence score of a character* > T_1) **AND** (*number of recognized characters* ≥ T_2), then it represents the frame with SC, here T_1 is a real-number in (0, 1.0) and T_2 is a positive integer. The absence (resp. presence) of SC in the candidate RS is used to label as replay (resp. live) frame. Shown in Fig. 1 is the illustration of various phases of the proposed SC detection process. For implementation of this work tesseract OCR method is used [26].

III. PERFORMANCE EVALUATION

Performance of the proposed system is evaluated on a video dataset consisting of 22 real-world sports videos. Objective metrics such as precision, recall, accuracy, and error rate are used for performance evaluation. The GUI of the implementation can be downloaded via [27].

A. Dataset

For performance evaluation, a dataset consisting of 22 real-world sports videos of a total duration of 10 hours is created. Each video in the dataset has a frame resolution of 640 x 480 pixels and a frame rate of 25 fps. Videos belong to four sports categories, i.e., *Cricket*, *Tennis*, *Baseball and Basketball*. The dataset consists of videos from five major broadcasters namely *ESPN*, *Ten Sports*, *Sky Sports*, *Fox Sports*, and *Euro Sports*. The experimental results are provided on the basis of system parameters that are set to α=3, *p*=2.5, *L*=5, T_1=0.6, T_2=5, N_{GT} = 10 N_{RL} = 50, and N_{RU} = 500.

The size of top hat filter α=3 is set to preserve the effectiveness of illumination adjustment and shape is set to *disk* for faster processing. For running average computation, window length *L*=5 is set to decrease the computational cost. For SC detection, threshold T_2 for number of recognized characters is set to 5 (i.e. T_2=5) because the minimum number of characters in SCs usually lie in the range of 5 to 6. If a character is recognized with more than 60% confidence (i.e. T_1 = 0.6) then it is recognized as a character. It was observed from the dataset that on average a GT consists of 10 frames, and minimum and maximum replay duration lie in the range of 2 to 20 seconds at 25 fps. Therefore N_{GT} = 10, N_{RL} = 50, and N_{RU}= 500 are used for experiments.

B. Experimental Results

Effectiveness of the proposed system is evaluated by detecting replay segments and highlight generation for each video in the dataset. The detection performed by the proposed system for each video is shown in Table I. From Table I it can be observed that the proposed system performs best for cricket, tennis, and basketball and for baseball the results are appreciable. The slight variation in baseball results can be attributed to the fact that baseball videos used, were recorded under lights that caused uneven illumination. The videos captured in better lighting conditions (under sunlight) resulted in superior detection performance. It is worth mentioning that the score-caption detection stage improves the overall performance of the system at the cost of relatively higher computational requirement.

In our second experiment, performance of the proposed system has been evaluated using receiver operating characteristic (ROC) curve analysis. Shown in Fig. 3 are the ROC curves of the proposed system for videos of four sports types. From the results it can be observed that the proposed method is very effective in terms of classifying the replay and live video frames.

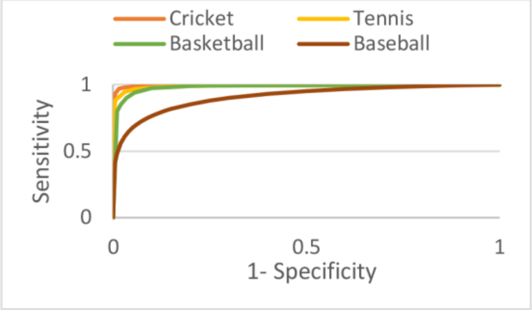

Fig. 3. ROC curves of sports videos.

In our last experiment, performance of the proposed system is compared with existing replay detection systems [1], [10], [12], [14], [16], [17], [21]-[23]. To this end, *precision* and *recall* are used for performance evaluation. Details of datasets used by each research group is provided in Table II. Shown in Table. II is the performance of the selected and proposed systems when tested on their respective datasets. It can be observed from Table. II that the proposed system achieves superior detection performance in terms of precision and recall when compared with existing state-of-the-art. In addition, effectiveness of the proposed system on four sports categories indicates that the proposed method is independent of underlying video type.

IV. CONCLUSION

In this paper, we propose a computationally efficient hybrid method for automatic sports highlights generation. The proposed method exploits the fact that a replay segment is sand-witched in gradual transitions and absence of score-caption in a replay segment. The proposed method is robust to broadcasters variation, sports category, score-caption design, camera variations, replay speed, and logo design, size, and placement. The proposed algorithm does not rely on logo template recognition for replay detection, which makes it computationally efficient. Effectiveness of the proposed method is evaluated on a diverse set of real-world videos. Experimental results indicate that the proposed system achieves average detection accuracy rate > 94%. It has been observed that under severe uneven illumination, performance of the proposed system degrades marginally. Currently, we are investigating performance of the proposed system on a bigger and more diverse dataset.

TABLE I

REPLAY DETECTION RESULTS FOR CRICKET, TENNIS, BASEBALL AND BASKETBALL.

Video Type	No. of frames	GT Start	GT End	True Positive	True Negative	False Positive	False Negative	Precision Rate	Recall Rate	Accuracy Rate	Error Rate
						Cricket					
Crick1	316	4	312	292	22	0	02	100%	99.31%	99.36%	0.64%
Crick2	320	16	318	292	25	02	02	99.31%	99.31%	99.06%	0.94%
Crick3	731	71	658	420	294	0	17	100%	96.11%	97.67%	2.33%
Average								**99.77%**	**98.24%**	**98.70%**	**1.30%**
						Tennis					
Tennis1	728	409	555	140	583	0	05	100%	96.55%	99.32%	0.68%
Tennis2	979	311	975	342	592	41	04	89.29%	98.84%	95.40%	4.60%
Tennis3	480	236	476	226	249	0	05	100%	97.83%	98.95%	1.05%
Average								**96.43%**	**97.74%**	**97.89%**	**2.11%**
						Baseball					
Base1	1053	118	1027	322	610	100	21	76.30%	93.87%	88.50%	11.50%
Base2	903	2	736	367	391	123	22	74.89%	94.34%	83.94%	16.06%
Base3	730	6	724	198	409	51	72	79.52%	73.34%	83.15%	16.85%
Average								**76.90%**	**87.18%**	**85.19%**	**14.80%**
						Basketball					
Basket1	627	143	584	266	349	10	02	96.37%	99.25%	98.09%	1.91%
Basket2	230	48	223	134	82	0	14	100%	90.54%	93.92%	6.08%
Basket3	356	52	321	211	139	0	6	100%	97.23%	98.31%	1.69%
Average								**98.79%**	**95.67%**	**96.78%**	**3.22%**

TABLE II

PERFORMANCE COMPARISON WITH EXISTING STATE-OF-THE-ART.

Techniques	**Non-Standard Dataset Details**						**Precision Rate**	**Recall Rate**
	Length (hours)	Format	Frame Rate	Resolution	No. of Videos	Sports Category		
Ekin et al. [1]	13	MPEG-1	30 fps	352 x 240	17	01	85.2%	80%
Pan et al. [10]	27	MPEG-2	25 fps	320 x 240	14	02	Not Used	94.6%
Zawba et al.[12]	02	AVI	30 fps	Not specified	05	01	81.15%	95.7%
Chang et al. [14]	18	Not specified	Not specified	Not specified	06	01	61.25%	77%
Nyugen et al. [16]	2:15	Not specified	Not specified	Not specified	03	01	94.6%	95.8%
Wang et al. [17]	2:30	Not specified	Not specified	Not specified	08	02	61.2%	74.77%
Xu et al. [21]	03	X264	30 fps	320 x 240	04	01	80.2%	81.1%
Eldib et al. [22]	06	Not specified	Not specified	Not specified	10	01	55.8%	80.7%
Chen et al. [23]	25	MPEG-2	30 fps	480x352	10	01	90%	92.8%
Proposed System	10	AVI	25 fps	640 x 480	22	04	98.8%	95.7%

REFERENCES

[1] A. Ekin, A.M. Tekalp, R. Mehrotra, “Automatic soccer video analysis and summarization,” *IEEE Trans. Image Process.*, vol. 12, no. 7, pp. 796–807, 2003.

[2] C. Li, Y.T. Wu, S.S. Yu, and T. Chen, "Motion-focusing key frame extraction and video summarization for lane surveillance system," in *Proc. 16th ICIP*, 2009, pp. 4329-4332.

[3] D. Anderson, R.H. Luke, J.M. Keller, M. Skubic, M. Rantz, and M. Aud, "Linguistic summarization of video for fall detection using voxel person and fuzzy logic," *Comp. Vision and Image Understand*., vol. 113, no. 1 pp.80-89, 2009.

[4] R.W. Lienhart, "Dynamic video summarization of home video," Electronic Imaging. International Society for Optics and Photonics, 1999.

[5] W.N. Lie, and C.M. Lai, "News video summarization based on spatial and motion feature analysis," *Advances in Multimedia Information Processing*, pp. 246-255, 2005.

[6] H.W. Chen, J.H. Kuo, W.T. Chu, and J.L. Wu, "Action movies segmentation and summarization based on tempo analysis," in *Proc. 6th ACM SIGMM Int. Workshop on Multimedia Inf. Retrieval*, 2004, pp. 251-258.

[7] B. Li, H. Pan, and I. Sezan, “A general framework for sports video summarization with its application to soccer,” in *Proc. ICASSP*, 2003, pp. 169–172.

[8] M.H. Hung and C.H. Hsieh, “Event Detection of Broadcast Baseball Videos” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 12, pp. 1713-1726, 2008.

[9] M. Tavassolipour, M. Karimian, and S. Kasaei, "Event detection and summarization in soccer videos using bayesian network and copula," *IEEE Trans. Circuits Syst. Video Technol.*, vol.24, no. 2, pp.291-304, 2014.

[10] H. Pan P.V. Beek, and M.I. Sezan, “Detection of slow-motion replay segments in sports video for highlights generation,” in *Proc. ICASSP*, 2001, pp.1649-1652.

[11] H. Pan, B. Lee, M.I. Sezan, “Automatic detection of replay segments in broadcast sports programs by detection of logos in scene transitions,” in *Proc. ICASSP*, 2002, pp.3385-3388.

[12] H.M. Zawbaa, N. El-Bendary, A.E. Hassanien, and T.H. Kim, "Machine learning-based soccer video summarization system,” *In Multimedia, Comp. Graphics and Broadcasting*, pp. 19-28, 2011.

[13] D. Tjondronegoro, Y. Chen, and B. Pham, “Highlights for more complete sports video summarization,” *IEEE Trans. Multimedia*, vol. 11 no. 4, pp. 22–37, 2004.

[14] P. Chang, M. Han, and Y. Gong, “Extract highlights from baseball game video with hidden Markov models,” in *Proc. ICIP*, 2002, pp.609–612.

[15] V. Kobla, D. DeMenthon, D. Doermann, “Detection of slow-motion replay sequences for identifying sports videos,” in *Proc. IEEE 3rd Workshop on MSP*, 1999, pp.135-140.

[16] N. Nguyen, and A. Yoshitaka, "Shot type and replay detection for soccer video parsing." in *Proc. IEEE ISM*, 2012, pp. 344-347.

[17] Wang, Lei, Xu Liu, Steve Lin, Guangyou Xu, and Heung-Yeung Shum. "Generic slow-motion replay detection in sports video." in *Proc. ICIP*, 2004. pp. 1585-1588.

[18] N. Babaguchi, K. Ohara, and T. Ogura, “Learning personal preference from viewer’s operations for browsing and its application to baseball video retrieval and summarization,” *IEEE Trans. Multimedia*, vol. 9, no. 5, pp. 1016-1025, 2007.

[19] H. Bai, W. Hu, T. Wang, X. Tong, C. Liu, and Y. Zhang. "A novel sports video logo detector based on motion analysis." *Neural Inf. Process.*, SpringerLink, pp. 448-457, 2006.

[20] L.Y. Duan, M. Xu, Q. Tian, and C.S. Xu, “Mean shift based video segment representation and applications to replay detection.” in *Proc. ICASSP*, 2004 pp.709–712.

[21] W. Xu, and Y. Yi, "A robust replay detection algorithm for soccer video," *IEEE Signal Process. Lett.*, vol. 18, no. 9, Sep. 2011, pp.509-512.

[22] M.Y. Eldib, B. Zaid, H. M. Zawbaa, M. El-Zahar, and M. El-Saban, "Soccer video summarization using enhanced logo detection," in *Proc. ICIP*, 2009, pp. 4345-4348.

[23] C.M. Chen, and L.H. Chen, "A novel method for slow motion replay detection in broadcast basketball video," *Multimedia Tools and App.*, vol. 74, no. 21, pp. 9573-9593, 2015.

[24] H.J. Zhang, A. Kankanhalli, and S.W. Smoliar. "Automatic partitioning of full-motion video." *Multimedia Sys.*, vol .1 no.1, pp.10-28, 1993.

[25] M. Zeng, J. Li, and Z. Peng. "The design of top-hat morphological filter and application to infrared target detection." Infrared Phys. & Tech., vol. 48, no. 1, pp. 67-76, 2006.

[26] R. Smith. An Overview of the Tesseract OCR Engine, In Proc. 9th International Conference on Document Analysis and Recognition (ICDAR), ’07, vol. 2, pp. 629-633, 2007

[27] Demo Link, [online]. Available: http://www-personal.engin.umd.umich.edu/~hafiz/projs/avs.htm.