Selection of keyframes for video colourization using steerable filtering

SOMDYUTI PAUL^{*}, SAUMIK BHATTACHARYA and SUMANA GUPTA

Department of Electrical Engineering, Indian Institute of Technology Kanpur, Kanpur 208016, India
e-mail: somdyutip@gmail.com; saumik@iitk.ac.in; sumana@iitk.ac.in

MS received 9 September 2015; revised 1 February 2017; accepted 11 February 2017; published online 8 September 2017

Abstract. The appropriate selection of distinctive keyframes to represent the salient contents of a video is a critical task in video processing applications that rely on content analysis or information retrieval. Although many of the existing keyframe selection techniques perform satisfactorily in capturing salient visual contents, they often fail to adequately highlight the changes in visual information brought about by motion of objects between frames. In this paper, we propose a technique for keyframe selection by formulating the dissimilarity between the frames of a video shot in terms of the change in orientations that the corresponding objects of the two frames have undergone due to motion. This is accomplished by steerable filtering of the frames in order to extract the information about the local orientation of pixels within each frame. The frame to frame dissimilarity is adaptively thresholded over a group of frames in order to select the keyframes. In essence, keyframes are selected at the temporal instances where the change in orientation attains local maxima. Our keyframe selection methodology is specifically relevant to video colourization due to the fact that the keyframes that are to be employed for colourization must be chosen such that they capture all orientational changes effectively, while ensuring adequate content coverage.

Keywords. Steerable filtering; dominant orientation; keyframe selection; adaptive thresholding.

1. Introduction

In many applications of video processing, it is a requisite to have one or more representative frames, termed as ‘keyframes’, which can effectively capture the intrinsic nature of the video shot. Representation of a video with the help of keyframes is of paramount importance in the task of video summarization, and allows for efficient video indexing, browsing and retrieval. Keyframes can be used to perform query-based searches to retrieve a video of interest from the world wide web. The use of keyframes to represent a video has also been employed for the purpose of video coding in order to improve compression efficiency.

The selection strategy used to choose keyframes from a given video shot depends on the intended application to a great extent. The criteria that the selected set of keyframes must meet also vary from one application to another. Moreover, for some applications, it suffices to select a predetermined number of frames as keyframes [1–4] (e.g., in video summarization often a single frame is used to represent an entire shot), whereas other applications require the number of keyframes selected to be dependent on the content and nature of video [5, 6]. One such application is video colourization, where the objective is to produce a

coloured video output from a given greyscale video. Many colourization techniques that have been developed so far allow the user to manually mark the desired colours as scribbles on the keyframes, which form a small subset of the entire set of frames. Such techniques seek to algorithmically propagate the scribble colours from the keyframes to the other frames [7–9]. However, although appropriate selection of keyframes is of immense importance to ensure the accuracy and effectiveness of the colourization process, existing scribble-based video colourization algorithms do not adequately emphasize on the technique used for keyframe selection. Most of such schemes assume that keyframes are selected manually, which can lead to improper selection of keyframes with poor content coverage since it is difficult to identify possible keyframes purely on a subjective basis by simply observing a video stream. In addition, manual keyframe selection is likely to cause reasonable fatigue of the user for video shots of long duration. Many conventional video colourization techniques such as [7–9] employ motion vectors in some form to propagate colours of the scribbles from the keyframe to the subsequent frames. Hence, the keyframes that are to be employed for video colourization, in general, should capture all significant changes in object positions and orientations to allow correct estimation of motion vectors. This is essential since if the two frames between which the motion vectors are estimated vary to a great extent in terms

^{*}For correspondence

of object orientations, then there is a strong likelihood that the motion vector estimation would be erroneous, causing colour artifacts in the colourized video. At the same time, the keyframes to be used for the purpose of video colourization should possess minimum redundancy with other keyframes of the same shot to reduce the manual intervention required by the user, as well as maximum correlation with its neighbouring frames, to ensure effective colour propagation.

In this paper, we develop a new technique for keyframe selection with a special focus on the applicability of the technique for video colourization. The number of keyframes chosen by our approach is highly dependent on the video content, and selection of keyframes relies on a dissimilarity measure derived from the change in dominant orientations of the steerable filter response of the video frames. Using the proposed approach of keyframe selection, with the colourization algorithm of [19] we have obtained robust colourization results. The rest of the paper is organized as follows. Section 2 reviews the existing techniques of keyframe selection. Section 3 elaborates on the proposed approach for keyframe selection. In section 4, the results of keyframe selection for different test sequences are presented. Finally, conclusions are drawn in section 5.

2. Related work

Many of the existing techniques for keyframe selection rely on a single keyframe to represent a given video sequence [1, 2]. Such techniques are useful for video browsing, and query-based video searches on the web. The technique of [1] computes the discrete cosine transform (DCT) feature vectors for each frame, and a cosine distance is used to form a similarity measure; thereafter, discriminative keyframes are found using linear discriminant analysis (LDA). The method of [2] combines the measures of motion activity, spatial activity, presence of faces and skin colour to produce an informative keyframe representing an entire video. Keyframe selection techniques based on clustering were proposed in [3, 4], which attempt to partition a given video shot into a fixed number of clusters, and select one keyframe each from the significant clusters. The keyframe selection strategy of [10] formulates keyframe selection as finding the predetermined number of frames that minimizes the temporal distortion introduced by summarization using a distortion metric based on principal component analysis (PCA). A keyframe selection technique based on SIFT keypoint matching was proposed in [6], which utilizes local features of a frame instead of global features to find the best representative frames of a video that have minimum redundancy amongst themselves and maximum content coverage. In the work of [5], motion analysis is used to select

keyframes. The author notes that in many video shots, stillness can be used to identify keyframes, and hence keyframes are selected at local minima of motion, using a motion metric computed using optical flow. In [11], a normalized cut criterion was introduced to select keyframes by partitioning the graph constructed using the shots of a video as nodes into subgraphs. There exist other keyframe selection techniques such as that of [12, 13] that exploit the information produced by video encoders like MPEG and H.264 to select keyframes that can represent the salient contents of the video well, and thereby improve the compression efficiency.

Many of the keyframe selection algorithms utilize the information of the colour channels of the video to compute the features used for demarcating the keyframes from the other frames [1, 3, 4]. However, our interest lies in selecting keyframes to be employed for a general application like video colourization. It is apparent that any video colourization algorithm takes a greyscale video as an input for which the colour channel information is not available, and has to be determined by the colourization algorithm itself. Similarly, for performing colourization of uncompressed video streams, neither do we possess the encoder information that can be exploited for keyframe selection as in the methods of [11–13]. This calls for an effective technique that can select keyframes without relying on such information. Moreover, keyframe selection techniques that primarily seek to capture the salient contents of a video shot for the purpose of summarization mostly detect keyframes only at those temporal instances where there is a significant change in the video content with respect to previous frames, such as the appearance of a new object in the scene. Such approaches generally do not detect a new keyframe when an object that was already present in the scene undergoes a large change in position as well as orientation due to motion. On the contrary, keyframes that are to be utilized for applications such as colourization should be generalized in the sense that they must highlight the changes due to motion, namely the positional and orientational changes of the moving objects, in addition to effectively capturing the salient visual contents of the video shot. This is essential to ensure correct colour propagation from the keyframe to the subsequent frames. The technique of [5] takes motion into account by incorporating the local minima of optical flow for keyframe selection. Nevertheless, optical flow gives incorrect estimate of motion in case of changes in ambient illumination amongst video frames, and also does not work for all types of motion such as zoom and affine motion. Unlike existing techniques, the keyframe selection methodology developed in this paper fulfils the aforementioned criteria for a generalized keyframe and also does not require any additional information such as that provided by colour channels or video encoders.

3. Proposed technique

The proposed approach for keyframe selection is applied on videos that are pre-segmented into shots using any existing shot detection scheme such as [14, 15]. The proposed algorithm detects the keyframes within each shot. As already discussed, keyframe selection methodology should ensure that the keyframes selected possess the property of having minimum redundancy with other keyframes of the same shot, and maximum correlation with their neighbouring frames.

Our approach explicitly assumes that the first and last frames of a video shot are always taken as keyframes, since there is an abrupt change in video content at these frames due to shot change. The frames of the video shot are subjected to steerable filtering in order to acquire the information about local orientation of spatial discontinuities within each frame. Without loss of generality, we assume that frames that have local maxima in terms of change in object orientations with respect to the nearest keyframe are the frames where the visual content have undergone an appreciable change due to motion and hence such frames should be selected as keyframes. Also, in case a new object enters the scene, the filter response will be drastically different for the pixels corresponding to the new object, and the local orientations of all such pixels will also be different, thereby accounting for the change. Thus, we formulate a dissimilarity measure based on the changes in orientation of steerable filter responses with respect to the nearest keyframe. Thereafter, keyframes are selected by adaptively thresholding this dissimilarity measure over a sliding window comprising a group of frames.

3.1 Steerable filtering

Steerable filters have found a wide range of applications in the analysis of images and video sequences such as the analysis of texture [17] and motion [16]. Particularly, steerable filters are a useful tool for the analysis of oriented structures present in an image or video. The application of the same filter rotated to different orientations on an image/video can effectively highlight the local orientation of the contours present in the image/video. The architecture for the construction of steerable filters was developed in [18], which establishes the fact that a filter can be steered to any arbitrary orientation by linearly combining the outputs of a set of basis filters with the help of direction-dependent steering coefficients provided the conditions imposed by the steering theorem are satisfied. A steerable filter may be expressed as follows:

$$f^\theta(x, y) = \sum_{j=1}^M k_j(\theta) f^{\theta_j}(x, y). \quad (1)$$

Here, $f^\theta(x, y)$ is the 2D steerable filter, which is steered to the orientation given by angle θ , $f^{\theta_j}(x, y)$ are the basis

functions and $k_j(\theta)$ are the orientation-dependent steering coefficients. In the polar coordinates, the function $f(x, y)$ can be expanded in Fourier series as follows:

$$f(r, \phi) = \sum_{n=-N}^N a_n(r) e^{in\phi}. \quad (2)$$

Here, $r = \sqrt{x^2 + y^2}$ and $\phi = \arg(x, y)$.

Let T be the number of non-zero coefficients $a_n(r)$ in Eq. (2). According to the steering theorem, Eq. (1) holds if and only if

$M \geq T$ and

$$\begin{pmatrix} 1 \\ e^{i\theta} \\ \vdots \\ e^{iN\theta} \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{i\theta_1} & \cdots & e^{i\theta_M} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{iN\theta_1} & \cdots & e^{iN\theta_M} \end{pmatrix} \begin{pmatrix} k_1(\theta) \\ k_2(\theta) \\ \vdots \\ k_M(\theta) \end{pmatrix}. \quad (3)$$

This implies that the minimum number of basis functions required to steer a given function $f(x, y)$ is T , and the steering coefficients must be the solutions of the matrix equation given by Eq. (3). If, for any n , $a_n(r) = 0$, then the n^{th} row of the matrix on either side in Eq. (3) is omitted. The basis function orientations θ_j must be chosen to ensure that the columns of the matrix in Eq. (3) are linearly independent. In this work, we take the second derivative of Gaussian G_2 and its Hilbert transform H_2 to constitute a complex quadrature pair of steerable filters; G_2 forms the real part and H_2 forms the imaginary part of the filter response [18]. The basis filters for G_2 and H_2 and their corresponding filter responses are illustrated by figures 1 and 2, respectively, which distinctly reveal that each basis selectively accentuates some edges depending on their orientations.

The input to the keyframe selection algorithm is a video shot $\mathbf{V}$ of size $P \times Q \times T$ where each frame is of size $P \times Q$, and T is the number of frames in the shot. Any pixel in a frame located at (x, y) is denoted by $\mathbf{p}$, where x and y are the spatial coordinates. The steerable filter is assumed to have been steered to O different orientations, and the angle corresponding to the j^{th} orientation is given by θ_j . Each frame of the video shot is convolved with the steerable filter steered to O different orientations. Then, the complex steerable filter response for the t^{th} frame at pixel $\mathbf{p}$ is given by

$$Z'(\mathbf{p}) = [Z'_1(\mathbf{p}) \ Z'_2(\mathbf{p}) \cdots Z'_j(\mathbf{p}) \cdots Z'_O(\mathbf{p})]. \quad (4)$$

The steerable filter response for a frame highlights the relative distribution of spatial discontinuities within the frame, and their orientations. This information is effectively exploited to derive a measure of frame to frame dissimilarity, as will be discussed further in the following sections.

Figure 1. Basis filters of G_2 and their responses: top row shows the three basis filters, while the bottom row shows the corresponding filter responses for the 33rd frame of the *cartoon* sequence.

Figure 2. Basis filters of H_2 and their responses: Top row shows the four basis filters, while the bottom row shows the corresponding filter responses for the 33rd frame of the *cartoon* sequence.

3.2 Orientation analysis

As an object in motion changes its position from frame to frame, the location and orientation of the edges of the corresponding object will also change from one frame to another. The rate of this change depends on the rapidity of motion present in the video sequence under consideration. This change is reflected in the response of steerable filters. Specifically, on applying the steerable filter steered to a particular orientation on a frame, all those edges whose orientation closely matches with the filter's orientation will produce a strong response. Thus, if the filter is steered to a sufficiently large number of orientations, every edge will have a strong response along some orientation. The orientation for which the response is maximum out of all orientations is referred to as the dominant orientation. In our approach, the change in dominant orientation of edges from one frame to another is used to model the dissimilarity between the two frames. For this purpose, we record the angular value of the dominant orientation at every

pixel to form an orientation matrix for every frame of the video shot as given by

$$\Omega_t(\mathbf{p}) = \left\{ \theta_j \mid |Z_j^t(\mathbf{p})| > |Z_k^t(\mathbf{p})|, \forall k \in \{1 \dots O\}, k \neq j \right\}. \quad (5)$$

Here, $\Omega_t(\mathbf{p})$ is the orientation matrix entry corresponding to the t^{th} frame at pixel $\mathbf{p}$, and $|Z_j^t(\mathbf{p})|$ is the absolute value of the complex response of the filter steered to the j^{th} orientation at pixel $\mathbf{p}$.

3.3 Dissimilarity measure

As already discussed, the degree of dissimilarity between any two frames due to motion of objects is reflected by the pixel to pixel difference of dominant angular orientation between the two frames. The objective of the keyframe

selection algorithm is to select keyframes at the temporal instances where this dissimilarity attains local maxima with respect to the nearest keyframe. For any frame, the nearest keyframe that lies behind it in time is referred to as its backward keyframe, and the nearest keyframe that lies ahead of it in time is referred to as its forward keyframe. The dissimilarity measures for a frame are computed with respect to the immediate backward and forward keyframes.

For any particular frame at temporal index t , let the immediate backward keyframe be K_b , and the immediate forward keyframe be K_f . Then the dissimilarity measures for the frame at t are given as

$$D_b(t) = \frac{1}{P \times Q} \sum_{\mathbf{p}} |\Omega_{K_b}(\mathbf{p}) - \Omega_t(\mathbf{p})| / 360, \quad (6)$$

$$D_f(t) = \frac{1}{P \times Q} \sum_{\mathbf{p}} |\Omega_{K_f}(\mathbf{p}) - \Omega_t(\mathbf{p})| / 360.$$

Here, $D_b(t)$ is the backward dissimilarity measure computed with respect to K_b , and $D_f(t)$ is a forward dissimilarity measure computed with respect to K_f . $D_b(t)$ and $D_f(t)$ give us an estimate of the overall change in object orientations between the t^{th} frame and its immediate backward and forward keyframe, respectively.

3.4 Keyframe selection by adaptive thresholding

Since our algorithm explicitly assumes that the first and last frames of a shot are taken as keyframes, initially, we have $K_b = 1$ and $K_f = T$. Hence, the initial dissimilarity measures are computed with respect to the first and last frames. As already pointed out, keyframes are selected at the local maxima of the dissimilarity measures $D_b(t)$ and $D_f(t)$. This is accomplished by adaptively thresholding the dissimilarity measures over a sliding window of size W . The thresholds are computed using the distribution of the dissimilarity measures over the sliding window as follows:

$$\eta_b = \frac{\max_j \{D_b(t) D_b(t+1) \cdots D_b(t+j) \cdots D_b(t+(W-1))\}}{c \times \frac{1}{W} \sum_{j=0}^{W-1} D_b(t+j)},$$

$$\eta_f = \frac{\max_j \{D_f(t) D_f(t-1) \cdots D_f(t-j) \cdots D_f(t-(W-1))\}}{c \times \frac{1}{W} \sum_{j=0}^{W-1} D_f(t-j)}. \quad (7)$$

In the above equation, the sliding window starts from frame t , and extends over the next $W-1$ frames. Here, c is a constant, which has been determined heuristically. New keyframes are selected whenever $D_b(t)$ or $D_f(t)$ exceeds the thresholds η_b or η_f , respectively. In this formulation, the average value, as well as the maximum value, of the dissimilarity measure dictates the selection of keyframes within the window of length W . If the maximum value of the dissimilarity measure is much larger than its average

over the window, it indicates that the most of the frames within this window are similar, although a few frames exist that are sufficiently different from the nearest keyframes and are the prospective keyframe candidates. Due to the low average value of the dissimilarity measure, the threshold is boosted in this case, and hence only the most distinctive frames succeed in surmounting the threshold to get selected as keyframes. On the other hand, when the average value, as well as the maximum value, of the dissimilarity measure is high within a window, it serves to indicate that visual content is undergoing rapid change, and hence the threshold is lowered by the high average value so that more keyframes are selected over this window to reflect the change adequately.

Algorithm 1: Key frame selection in a video shot

Result: Display the keyframe numbers

Initialize

$K_b = 1, K_f = T$, Sliding window size W ;

$j_b = K_b + 1$;

$j_f = K_f - 1$;

Compute $D_b(K_b), D_b(K_b + 1) \cdots D_b(K_b + (W - 1))$;

Compute $D_f(K_f), D_f(K_f - 1) \cdots D_f(K_f - (W - 1))$;

η_b : computed over K_b to $K_b + (W - 1)$;

η_f : computed over K_f to $K_f - (W - 1)$;

while $j_f > j_b$ **do**

if $D_b(j_b) > \eta_b$ **then**

$K_b = j_b$;

 print K_b ;

 update $D_b(K_b), D_b(K_b + 1) \cdots D_b(K_b + (W - 1))$;

 recompute η_b

end

if $D_f(j_f) > \eta_f$ **then**

$K_f = j_f$;

 print K_f ;

 update $D_f(K_f), D_f(K_f - 1) \cdots D_f(K_f - (W - 1))$;

 recompute η_f

end

$j_b = j_b + 1$;

$j_f = j_f - 1$;

end

Starting with the first and last frames as initial keyframes, we compare $D_b(t)$ and $D_f(t)$ with the thresholds η_b and η_f computed over the corresponding sliding windows. Frames for which $D_b(t)$ exceeds η_b or $D_f(t)$ exceeds η_f are selected as new keyframes. The window slides to the position of the new keyframe. The dissimilarity measure and the corresponding threshold are recomputed over the new window, and further comparison is carried out in the same manner. This is illustrated with the help of figure 3, which shows the frames of a shot comprising T frames. The first frame F_1 and the last frame F_T are the default backward and forward keyframes, which are labelled in red. The initial positions of the sliding windows at the two ends of the shot are shown by the boxes outlined in blue. While

Figure 3. Keyframe selection using sliding windows for comparison of dissimilarity measures in forward and backward directions.

Table 1. Number of keyframes selected for test videos.

Video sequence	Total no. of frames	No. of keyframes	Percentage of keyframes (%)
<i>Akiyo</i>	242	2	0.83
<i>Coastguard</i>	300	6	2
<i>Cartoon</i>	92	4	4.34
<i>Flower</i>	250	44	17.60
<i>Skiing</i>	59	13	22.03
<i>Soccer</i>	240	149	62.08

making the backward and forward comparisons with respect to frames F_1 and F_T , respectively, suppose frames F_{t_b} and F_{t_f} , labelled in purple, are selected as the new keyframes. Immediately, the windows on either sides slide to the positions of F_{t_b} and F_{t_f} as shown by the boxes outlined in green. In this way, the windows on either side slide towards each other as shown by direction of the arrows. The process terminates when the windows cross each other. The keyframe selection process is summarized with the help of the pseudo-code given by algorithm 1.

4. Experimental results

In this section the results obtained on applying the proposed technique for keyframe selection are presented. We have used nine tap steerable filters, constructed according to

tables 3–6 of Appendix H of [18], which give the tap coefficients of the 1D filter, which are used to construct the separable pair of quadrature filters, along with the steering coefficients of each basis. The filter is steered to 180 different orientations in the 2D space, which is found to be sufficient to capture any discernible change in object orientations between frames. The constant c is set to 10 in our work. The size of the sliding window is taken to be 10 frames. The percentages of keyframes obtained for six test video sequences are enlisted in table 1. It is evident that the percentage of keyframes varies widely from one sequence to another depending on the magnitude of motion and texture content of the sequence. For highly dynamic video sequences like *Soccer*, where the position and orientation of objects change rapidly between consecutive frames, the proposed technique produces a high percentage of keyframes. On the other hand, for a sequence like *Akiyo*, which is characterized by very small changes between consecutive frames due to the very low motion content, our algorithm selects a much smaller percentage of frames as keyframes. The *Flower* sequence also produces a reasonable percentage of keyframes, since it is rich in texture content, and even a small shift in the textured pattern from one frame to another produces a large change in terms of orientations. In contrast, the *cartoon* sequence, which has flat regions with low texture, gives rise to a very few keyframes, which are selected at points where the visual content undergoes a substantial change due to motion. For the *Coastguard* sequence, the upward tilt of the camera causes a rapid

Figure 4. Keyframes of the *Skiing* sequence.

Figure 5. Keyframes of the *cartoon* sequence.

change in visual content and hence more keyframes are selected from the region where the camera undergoes the upward tilt. Figure 4 shows the keyframes selected by our algorithm for the *Skiing* sequence. It can be seen that the changes in the orientations of the moving person are depicted well by the keyframes. Figure 5 illustrates the keyframes selected for the *cartoon* sequence, which capture the salient changes in the video sequence.

We provide the video sequences mentioned in table 1, and the corresponding keyframes selected by the proposed technique.¹ To demonstrate that the proposed keyframe selection technique is effective for video colourization, some of the colourization results obtained by the method of [19], which relies on the keyframes selected by the proposed technique to provide colour sources as scribbles, are also made available.²

5. Conclusion

In this paper, we have developed an effective technique for keyframe selection, which can produce distinctive keyframes from a video sequence by taking into account the inherent motion and texture content of the sequence. By effectively exploiting the information about the change in dominant orientation with respect to the nearest keyframe to derive a measure of dissimilarity for frames, it is possible to obtain an estimate of the positional and orientational changes that objects in motion undergo between frames; thus, keyframes can be selected at the temporal instances where this change is substantial. In this way, changes occurring in the existing objects of a scene due to motion can be incorporated into the keyframe selection process, rather than merely focussing on the changes in visual information such as the appearance of new objects in the scene. In contrast to many existing approaches, our approach does not rely on the information present in the colour channels of the video, or the information extracted from video encoders. This makes our approach suitable for all classes of video sequences.

The experimental results obtained on performing keyframe selection for a wide variety of test sequences

establish that the percentage of frames that are chosen as keyframes by our algorithm directly reflects the motion and texture pattern of the sequence. This factor makes our approach especially useful for selecting keyframes that are to be employed for video colourization. As we have emphasized earlier, the keyframes that are to be employed for colourization should highlight all the significant changes in the positions as well as orientations of objects in order to ensure accurate temporal propagation of colour, and the results of keyframe selection for the various test sequences satisfy these criteria well. Further, since the keyframes selected by our approach effectively incorporate the salient visual information within a video shot, they are also suitable for creating a summary of the shot for representative purposes.

References

- [1] Cooper M and Foote J 2005 Discriminative techniques for keyframe selection. In: *Proceedings of the IEEE International Conference on Multimedia and Expo*, pp. 502–506
- [2] Dirfaux F 2000 Keyframe selection to represent a video. In: *Proceedings of the International Conference on Image Processing*, pp. 275–278
- [3] Zhuang Y, Rui Y, Huang T S and Mehrotra S 1998 Adaptive key frame extraction using unsupervised clustering. In: *Proceedings of the International Conference on Image Processing*, pp. 866–870
- [4] Girgensohn A and Boreczky J 1999 Time-constrained keyframe selection technique. In: *Proceedings of the IEEE International Conference on Multimedia, Computing and Systems*, pp. 756–761
- [5] Wolf W 1996 Key frame selection by motion analysis. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1228–1231
- [6] Guan G, Wang Z, Lu S, Deng J D and Feng D D 2012 Keypoint based keyframe selection. *IEEE Trans. Circuits Syst. Video Technol.* 23: 729–734
- [7] Levin A, Lischinski D and Weiss Y 2004 Colorization using optimization. *ACM Trans. Graph.* 23: 689–694
- [8] Heu J H, Hyun D Y, Kim C S and Lee S U 2009 Image and video colorization based on prioritized source propagation. In: *Proceedings of the IEEE International Conference on Image Processing*, pp. 465–468

¹Keyframe selection: <http://goo.gl/lnsf2p>.

²Colourization results: <https://goo.gl/57dGhm>.

- [9] Sheng B, Sun H, Magnor M and Li P 2013 Video colorization using parallel optimization in feature space. *IEEE Trans. Circuits Syst. Video Technol.* 24: 407–417
- [10] Li Z, Schuster G M and Katsaggelos A K 2005 MINMAX optimal video summarization. *IEEE Trans. Circuits Syst. Video Technol.* 15: 1245–1256
- [11] Ngo C W, Ma Y F and Zhang H J 2005 Video summarization and scene detection by graph modeling. *IEEE Trans. Circuits Syst. Video Technol.* 15: 296–305
- [12] Usach-Molina P, Sastre J, Naranjo V, Vergara L and Munoz J M L 2010 Content-based dynamic threshold method for real-time keyframe selecting. *IEEE Trans. Circuits Syst. Video Technol.* 20: 982–993
- [13] Liu G and Zhao J 2010 Keyframe extraction from MPEG video stream. In: *Proceedings of the Third International Symposium on Information Processing*, pp. 423–427
- [14] Liu G, Wen X, Zheng W and He P 2009 Shot boundary detection and keyframe extraction based on scale invariant feature transform. In: *Proceedings of the Eighth IEEE International Conference on Computer and Information Science*, pp. 1126–1130
- [15] Cernekova Z, Pitas I and Nikou C 2005 Information theory-based shot cut/fade detection and video summarization. *IEEE Trans. Circuits Syst. Video Technol.* 16: 82–91
- [16] Zhen X and Shao L 2013 Spatio-temporal steerable pyramid for human action recognition. In: *Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, pp. 1–6
- [17] Montoya-Zegarra J A, Leite N J and Torres R D S 2007 Rotation-invariant and scale-invariant steerable pyramid decomposition for texture image retrieval. In: *Proceedings of the XX Brazilian Symposium on Computer Graphics and Image Processing*, pp. 121–128
- [18] Freeman W T and Adelson E H 1991 The design and use of steerable filters. *IEEE Trans. Pattern Anal. Mach. Intell.* 13: 891–906
- [19] Paul S, Bhattacharya S and Gupta S 2016 Spatiotemporal colorization of video using 3D steerable filtering. *IEEE Trans. Circuits Syst. Video Technol.*, DOI [10.1109/TCSVT.2016.2539539](https://doi.org/10.1109/TCSVT.2016.2539539)