Received February 14, 2018, accepted April 5, 2018, date of publication April 9, 2018, date of current version May 2, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2825106

Dynamic Mode Decomposition Based Video Shot Detection

CHONGKE BI[1], YE YUAN[1], JIAWAN ZHANG[1], YUN SHI[2], YIQING XIANG[1], YUEHUAN WANG[1], AND RONGHUI ZHANG[3]

[1]School of Computer Software, Tianjin University, Tianjin 300350, China
[2]Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 10008, China
[3]Guangdong Key Laboratory of Intelligent Transportation System and Research Center of Intelligent Transportation System, School of Engineering, Sun Yat-sen University, Guangzhou 510275, China

Corresponding authors: Ye Yuan (yuanye1989@tju.edu.cn) and Jiawan Zhang (jwzhang@tju.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61702360 and Grant 51775565.

ABSTRACT Shot detection is widely used in video semantic analysis, video scene segmentation, and video retrieval. However, this is still a challenging task, due to the weak boundary and a sudden change in brightness or foreground objects. In this paper, we propose a new framework based on dynamic mode decomposition (DMD) for shot boundary detection. Because the DMD can extract several temporal foreground modes and one temporal background mode from video data, shot boundaries can be detected when the amplitude changes sharply. Here, the amplitude is the DMD coefficient to restore the video. The main idea behind the shot boundaries detection is finding the amplitude change of background mode. We can reduce error detection when the illumination changes sharply or the foreground object (or camera) moves very quickly. At the same time, our algorithm has a high detection accuracy, even the color changes are not obvious, the illumination changes slowly, or the foreground objects overlap. Meanwhile, a color space for DMD is selected for reducing false detection. Finally, the effectiveness of our method will be demonstrated through detecting the shot boundaries of the various content types of videos.

INDEX TERMS Dynamic mode decomposition, video shot detection, shot boundary.

I. INTRODUCTION

Most videos such as movies and documentaries, include acts, scenes, and shots. There are multiple scenes in an act, and a scene contains multiple shots [1]–[4]. The director usually uses the shot transitions to switch the shots more smoothly, so that the video connection will be more coherent [5]–[8]. Common shot transitions include: hard cuts, fades and dissolves.

1) Hard cut:

As shown in Figure 1(a), it is the most basic method of shot transitions.

2) Fade out:

As Figure 1 (b) shows, this shot transition process is very slow. The scenes and foreground objects in the shots are almost unchanged, only the brightness changes slowly. This could bring some difficulties to the shot detection.

3) Dissolve:

The images of the previous shot are overlapped on the images of the next shot, as Figure 1 (c) shows. It can be seen that the changes of brightness and color are not obvious, and the images in the two shots overlap. On the boundary of two shots, there will remain the visual features of both, the front and the next shot. It leads to error detection by using the video content including texture, color and shape.

There are many existing methods for shot detection. Some methods use pixel, pixel block and histogram comparison to detect shot boundaries. The others are based on video content and visual characteristics for shot detection including texture, color and shape. Moreover, there are many approaches to detect the shot boundary by cosine similarity, weighted edge information, machine learning and so on. Although the existing methods can be used for shot segmentation, there still remain many problems [9], [10]. When using pixel, pixel block and histogram comparison for shot detection, it has a good effect when the brightness of the shot boundary changes obviously. However, these methods are easy to fail. Because the illumination changes drastically at the non-shot boundary [11], [12]. When the object (or camera) moves fast, the correct rate of the algorithms will greatly reduce. Due to

FIGURE 1. Three shot transition methods: (a) hard cuts with no medium frames; (b) fades; (c) dissolves.

the overlap of two shots images, the shot detection methods based on video content, including texture, color and shape, have low accuracy [13].

Moreover, the recall and accuracy of processing fades and dissolves are very low by using cosine similarity, weighted edge information, machine learning and so on. Although there have been a lot of researches on shot detection [14]–[16], it is still a challenging task because there are two problems which lead to lower precision and lower recall.

Q1: In non-shots transitions, the illumination changes sharply, or the foreground object (or camera) moves faster. Existing algorithms may have error detection.

Q2: In shot transitions, the color difference in the different scene is too small (Hard cut), the illumination changes slowly (fade), or the foreground objects overlap together (dissolve). Existing algorithms may have missed detection.

According to the two questions above, we summarize the two conditions to be satisfied in the process of shot detection:

C1: In non-shots transitions, the feature weights used to detect shot boundaries should be very stable.

C2: In shot transitions, the feature weights used to detect shot boundaries change significantly and are easy to identify.

The existing methods are analyzing every image in the video data and extract the features (pixel, pixel blocks, texture, etc.) of a single image [17], [18]. Then, the features are applied to each frame of the video data. In this process, the temporal features of the video data are ignored. It will lead to Q1 and Q2 problems.

We try to find a way to satisfy the two conditions in the shot detection process. In this paper, we propose a new framework for shot boundary detection based on dynamic mode decomposition (DMD). Our method is to take video data as a sequence of matrix data, and extract the temporal features directly from the matrix data. After that, we use DMD to extract a temporal background mode and a series of temporal foreground modes in the video. The background (or foreground) of each frame can be restored by using the temporal background (or foreground) modes and its amplitude. The amplitude is the temporal feature weight of the background (or foreground) mode. For video data without shot boundaries, the temporal background mode is constant, and the amplitude of the temporal background mode is stable. This process satisfies the condition C1. On the contrary, the temporal background mode and amplitude have an acute change in a video data with shot boundary. Therefore, we use the temporal background mode amplitude to detect the shot boundary. Meanwhile, this shot detection process satisfies the condition C2. Meanwhile, a color space for DMD is selected for reducing false and missed detection.

A large number of experimental videos contain some movies and movie trailers [19]. The experimental results not only verify the effectiveness of our method, but also show that the average accuracy and recall rate of the proposed method are also very high in the existing methods. In particular, the major contributions of this paper are as follows:

1) We propose a new framework for shot boundary detection, which can detect different kinds of shot transition including hard cut, fade and dissolve instead of using a series of algorithms.

2) Our algorithm can solve the problem caused by the illumination changes sharply or the foreground object (or camera) moves very quickly. It has a high accuracy rate and recall rate.

The rest of this paper is organized as follows. Section II describes a brief survey on related work. Section III presents our framework. Section IV introduces the method for shot detection. Section V describes the usage of background mode amplitude for shot detection. Section VI introduces the color space used for DMD method. The experimental results and analysis are given in Section VII. Finally, we conclude this paper and discuss possible future work in Section VIII.

II. RELATED WROK

In this section, we mainly introduce the definition of shot and shot boundary, the existing methods of shot detection and the related work of dynamic mode decomposition method.

A. SHOT BOUNDARY DETECTION ALGORITHMS

Since video shot detection is the foundation of almost all of the video content analysis, segmentation and retrieval, it is important to get the accurate shot boundaries results to do the following work [20]. Although it seems easy for human being to find the cuts of films, computers cannot always have all the cuts detected correctly. Possibly it is hardly capable to achieve the goal that divides the film to shots properly enough, unless there is powerful artificial intelligence. Most of the existed algorithms have good results on the hard cuts, but fail in configuring gradual transitions like fade out and dissolve. On the other hand, some algorithms can deal with one type of shot transitions, not adaptive to other types of gradual cuts or hard cuts.

All along, many scholars have done the work to shot transition detection [21]. The main idea to figure out the shot boundaries is that define the similarity/dissimilarity between two frames and detect the mutant frame as the boundary [22].

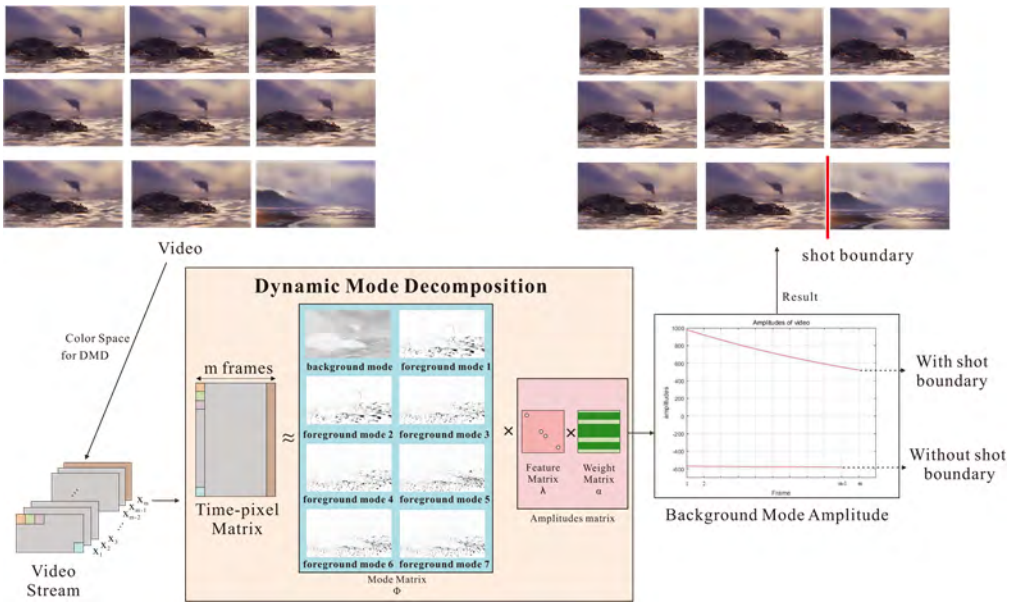


FIGURE 2. Framework of video shot detection, including three steps: noise reduction; feature extraction using DMD; and boundary detection using background mode from DMD.

Each shot detection method works on two steps to get the results. The first is scoring, finding a representative way to perform every frame feature, which can be used to distinguish the irrelevant content. The second is decision, evaluating the previous value using a threshold to get the cuts. Using the visual features to represent every frame to get the boundaries is the usual method to do cut detection. Early shot detection is mainly based on pixel [23], pixel block [24] and histogram comparison [25]. Color histogram is one of them, which is proposed by Mas and Fernandez [26]. Using color histogram differences to differentiate the abrupt shot boundaries and the temporal color variation to smooth shot boundaries. It shows good performance with the abrupt shot transition and the moving objects. However, for the soft shot transitions, it is needed to be improved or draw support from some other methods to get the ideal results. Moreover, the shot detection algorithms based on pixel or pixel block encounter some of the same issues. Obviously, the separate visual method cannot solve the shot detection problem perfectly.

Other algorithms are based on video content for shot detection, including texture [27], color [28] and shape [29]. Wang *et al.* [30] present a shot detection method based on texture features. This algorithm is based on the detection of video content. Although it can solve the noise effects caused by color mutation, this method cannot solve the problems caused by the fast motion of the object or camera in the video. The latest approaches detect the shot boundary by using cosine similarity [31], weighted edge information [32], machine learning, sift features and so on. Lin *et al.* [33] and Desobry *et al.* [34] proposed a support vector machine (SVM) based machine learning method for shot detection. This method is more efficient and robust to noise. Although with the development of machine learning, artificial intelligence can be used to detect cuts [35], it has not got the breakthrough progress. However, this method is not good for the shot transitions of fade and dissolve. Usually, each type of boundaries needs a special method to be detected, leading the problem to be more complicated with many mathematical formulas and manually set thresholds. An algorithm with university that can be applied to most of common boundaries is urgently needed.

B. DYNAMIC MODE DECOMPOSITION

The dynamic mode decomposition [36] can provide precise reconstructions of spatio-temporal coherent structures. It works on the nonlinear dynamical systems [37] in an equation-free, data-driven matrix way. DMD is such a mathematical method that is able to extract the content modes from the sequential data set, which has successfully been used in the fluid mechanics for the data analysis and discovery work [38], [39]. Since the video is made by the frames with time, it has been used to separate the background and foreground of the video and art image segmentation successfully [40]. It is deduced that DMD method can be used to process the video data sets. Every shot is the composition of a sequence of related frames. The shot boundaries can be detected with the modes calculated by DMD.

III. OUR FRAMEWORK

In this paper, we propose a new framework based on dynamic mode decomposition to detect shot boundaries from three different shot transitions including hard cuts, fades and dissolves. As Figure 2 shows that our framework is divided into three steps.

By selecting a color space for DMD, we can enlarge the temporal feature weight of the background (or foreground) mode, reduce the influence of noise (brightness, texture, etc.) on shot detection, and effectively reduce missed detection and error detection. Through experiments, we recommend to use the HSV color space.

Then, we can use DMD to decompose video data mode matrices and amplitude coefficient matrices. The mode matrices consist of a temporal background mode and several temporal foreground modes. The amplitude matrix contains the temporal feature matrix and the weight matrix. These three matrices can be used to recover the background and foreground images of each frame in the video.

Finally, the difference between the backgrounds of each frame image is determined by the amplitude. For a video data without shot boundaries, the temporal background mode is constant, and the amplitude of the temporal background mode is stable. However, if there is a shot boundary in the video, the temporal background mode and amplitude will have an acute change. Because the video background changes drastically at the shot boundary, we use the change rate of temporal background mode amplitude to detect the shot boundary. When the amplitude changes violently, the last frame is the shot boundary, the opposite is not.

IV. DYNAMIC MODE DECOMPOSITION BASED SHOT DETECTION

A. MERITS OF DYNAMIC MODE DECOMPOSITION

DMD was proposed by Schmid [19] to solve the problem in the field of fluid physics. The main advantage of DMD is used for mode analysis. The extracted mode can be used for data reduction. Especially for large-scale fluid simulation process, a coefficient matrix can be obtained to recover large-scale fluid data with low computational cost. Dynamic mode decomposition has not only been widely used in the physical field, but also applied in the fields of image and vision. The process of data reduction using DMD algorithm is actually decomposing the data into low rank and sparse part, which can be used for video and image feature extraction. Furthermore, in the process of background and foregrounds extraction, it is not necessary to adjust threshold using DMD algorithm. Therefore, the computational complexity is low compared to other algorithms such as RPCA, which needs complicated parameter adjustment.

B. DYNAMIC MODE DECOMPOSITION FOR SHOT DETECTION

The video data has a strong spatial and temporal correlation. Each video shot can be seen as a potentially complex nonlinear dynamic snapshot. Dynamic mode decomposition is a mathematical method, which focuses on the discovery of coherent spatio-temporal patterns of high dimensional data onto complex systems with temporal dynamics. Video shot detection is a challenging task, due to the complexity of shot transitions and mutation of brightness. The ability of the DMD is to discover and utilize the background changes in complex systems. This is the key to solve the shot detection.

The premise of the DMD algorithm is that the given data set must be a fixed interval of data. We define a video stream is from X_1 to X_N which is a uniform sampling data at N frames, and the time interval is Δt . Temporal representation of video data is

$$\mathbf{X}_1^N = [x_1\ x_2\ x_3\ \cdots\ x_N] \tag{1}$$

x_i $(1 \ll i \ll N)$ is an image. This image is a frame in the video. We assume that there is a linear map of the process, so we can use the Koopman operator $\mathbf{A}$ to map the data at the jth time to the data at the $j+1$th.

$$x_{i+1} = Ax_i \tag{2}$$

Since $\mathbf{X}_1^{n-1} = \mathbf{U}\sum\mathbf{V}^*\boldsymbol{S}$, we can deduce as follows

$$\mathbf{X}_2^N = \mathbf{A}\mathbf{X}_1^{N-1} \approx \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*\boldsymbol{S} \tag{3}$$

in which $\mathbf{U}$ is unitary ($\mathbf{U} \in C^{m\times l}$), $\boldsymbol{\Sigma}$ is diagonal ($\boldsymbol{\Sigma} \in C^{l\times l}$) and $\mathbf{V}$ is unitary ($\mathbf{V} \in C^{n-1\times l}$). Parameter l is chosen to minimize x rank. We can get following the matrix $\mathbf{S}$, which is determined from the matrices of $\mathbf{X}_1^{N-1}$ and $\mathbf{X}_2^N$ by minimizing the Frobenius norm of the difference between $\mathbf{A}\mathbf{X}_1^{N-1}$ and $\mathbf{X}_2^N$.

The matrix $\mathbf{S}$ is

$$\mathbf{S} \approx \mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{U}^*\mathbf{X}_2^N \tag{4}$$

Using the similarity transformation ($\mathbf{V}\boldsymbol{\Sigma}^{-1}$) to derive the matrix, which is similar to the matrix $\mathbf{S}$, and $\mathbf{S}$ is expressed as

$$\mathbf{S} \approx \mathbf{U}^*\mathbf{X}_2^N\mathbf{V}\boldsymbol{\Sigma}^{-1} \tag{5}$$

The basic idea of DMD algorithm is

$$\mathbf{A}\mathbf{X}_1^{n-1} = \mathbf{X}_2^n \approx \mathbf{X}_1^{n-1}\mathbf{S} \tag{6}$$

and then some of the eigenvalues of the matirx $\mathbf{S}$ approximate the eigenvalues of the Koopman operator $\mathbf{A}$, similar to calculation done in the Arnoldi algorithm to get the *Ritz values*. Consequently, $\mathbf{AU} \approx \mathbf{US}$. The DMD mode φ_j is:

$$\boldsymbol{\varphi}_\mathbf{j} = \mathbf{U}\boldsymbol{y}_j \tag{7}$$

In addition, transform the eigenvalue to Fourier mode to predict time dynamic:

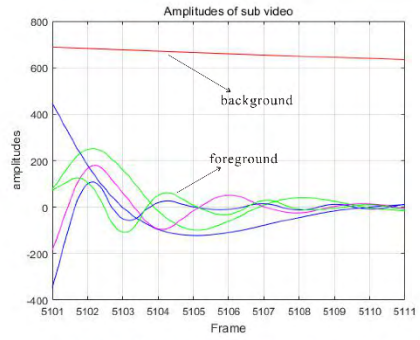
$$\boldsymbol{\omega}_\mathbf{j} = \frac{\ln(\boldsymbol{\mu}_j)}{\boldsymbol{\Delta t}} \tag{8}$$

The real part of ω_j corresponds to the growth or attenuation of the DMD basis function, and the imaginary part of ω_j corresponds to the oscillation of the DMD mode. Through $\mathbf{X}_{DMD}(t) = \mathbf{A}^t x_1$, we can reconstruct the video by

$$\mathbf{X}_{DMD}(\boldsymbol{t}) = \sum_{j=1}^{r}\boldsymbol{\alpha}_j\boldsymbol{\varphi}_j\boldsymbol{\mu}_j^{t-1} = \sum_{j=1}^{r}\boldsymbol{\alpha}_j\boldsymbol{\varphi}_j e^{\boldsymbol{\omega}_j t} \tag{9}$$

It is obvious that there will be a corresponding Fourier mode (ω_j) located near the far point of the complex space point, $||\omega_j|| \approx 0$, if the first frame of the video do not change

(a)Sub video with no cuts



(b)Amplitudes of the above sub video

FIGURE 3. Mode matrix, feature matrix and weight matrix decomposed from video data.

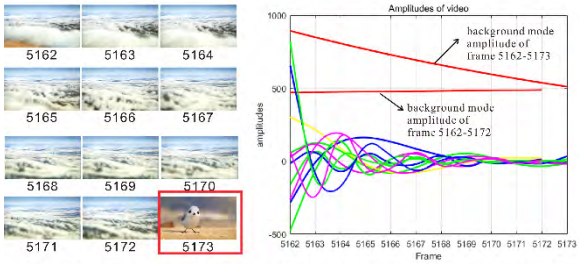
with time going on, or change very slowly. This fact become the key principle that DMD has the ability of extracting shots. The background mode represents a relatively stationary scene in a video, while the foreground modes represent a plurality of objects or scenes in relative motion in a video. As shown in Figure 3, the DMD can be used to calculate the sequential video data and get the mode matrix Φ, the feature matrix λ and the weight matrix α. Assuming that there exists ω_p, the p out of it belongs to $\{1, 2, 3, \ldots, l\}$ and makes bounded and is far from zero. We can get the background mode and foreground modes using Eqs. (10) and (11). Note that, in all figures of this paper, the background mode is represented by using a red line, while the foreground modes using lines with other colors. For our shot boundary detection method, the background mode is enough for fully extract all shot boundaries.

$$\mathbf{X}_{DMD}^{Background-mode} = \sum_{j=p}^{r} \alpha_j \varphi_j e^{\omega_j t} \tag{10}$$

$$\mathbf{X}_{\mathbf{DMD}}^{Foreground-mode} = \sum_{j\neq p}^{r} \alpha_j \varphi_j e^{\omega_j t} \tag{11}$$

Because the temporal feature matrix λ is calculated by the temporal background (foreground) mode, it is the features of the background and foreground in the video. The weight matrix represents the weight of the different modes in each frame of the image, including both the background and the foreground. The amplitudes corresponding to background and foreground modes are features. The variation of amplitude with time $\mathbf{A}_{amp}(t)$ is:

$$\mathbf{A}_{amp}(t) = \boldsymbol{\alpha} \times \boldsymbol{\lambda} = \sum_{j=1}^{r} \alpha_j \mu_j^{t-1} \tag{12}$$



(a)A video with a hard cut (b)Amplitudes of the above sub video

FIGURE 4. The amplitude lines (b) extracted from the video (a) using DMD. Red line: background mode amplitude; the other lines: foreground modes amplitudes.

V. SHOT DETECTION USING AMPLITUDE OF BACKGROUND MODE

DMD can extract the temporal background mode and temporal foreground modes in a video. The temporal background mode represents a relatively stationary background object or scene in a video for a period of time. And the temporal foreground modes represent relative motion, changing objects or scenes in the same video. In general, a video can be extracted to a temporal background mode and several temporal foreground modes. The amplitude is the temporal feature weight of the background (or foreground) mode, which represents the degree of change in the temporal background (or foreground) in a video. When the background (or foreground) changes violently, it is shown that the background (or foreground) object has changed acutely in the video. The temporal background mode and amplitude have an acute change in a video data with shot boundary. Therefore, the temporal amplitude calculated by DMD can be used for a shot detection.

Figure 4 shows a video containing twelve frame images in one shot. There is no shot boundary in the first eleven frames. The sea washed the shore in this shot. The sky is the background, which does not change, therefore the shot does not switch either. For this video using DMD factorization, the background mode amplitude with time can be seen:

1) It is shown that this model can capture the dominant content in the video. That is, the same background. This background is actually relative. Because when the shot movement or follow an object in the process, the object is the background. In Figure 4, this constant amplitude corresponds to the pattern of the background, which is the sky.

2) The next set of mode is the wave. With the movement of the camera, the wave is moving slowly, that is the foreground elements in the shot.

3) The other modes represent by other changes in the video sequence, such as small movements in the background.

However, there is a shot boundary in the last frame. In Figure 4 (b), we can see that the amplitude of 5162 to 5172 frames is very gentle. However the amplitude of 5162 to 5173 frames is very sharp. Hence, the 5173th frame is a shot boundary.

In order to accurately determine the shot boundary, we set an amplitude threshold to compare with the amplitude of temporal background mode. Through a large number of experiments, the recommended amplitude threshold value should be set as around 150. When the amplitude of the temporal background mode changes more than the amplitude threshold, there is a shot boundary. If the amplitude of the temporal background mode is less than the amplitude threshold, it shows that there is no shot boundary in the subvideo. Not only that, because the frame rate of the video is 24 fps, we take 24 as frames threshold for shot boundary detection. When a subvideo has reached 24 frames and the amplitude of the temporal background mode does not exceed the amplitude threshold (150), it means that there is no shot boundary in this subvideo. We use the twenty-fifth frame as the first frame to calculate the shot boundary detection. When the amplitude of temporal background mode changes beyond the amplitude threshold (150) during sub-video calculation, this means that the last frame is the shot boundary. We calculate and detect the shot boundary from the frame of shot boundary. This can greatly reduce the amount of computation and improve computational efficiency, to avoid calculating a large number of video frames.

VI. COLOR SPACE FOR DMD

Of course, using the background model to detect the shot boundary has the same problem of noise interference. In this section, we will introduce increase the temporal feature weights by using the different color space, thus reducing the impact of brightness on shot detection.

The traditional method is to compare the brightness difference between adjacent frames to determine the shot boundary. The brightness of the shot boundary will have a huge difference. But when the illumination changes, the precision rate of the algorithms will be greatly reduced. Therefore, the noise of traditional shot detection algorithms will occur to the brightness changes of two adjacent frames.

We find the shot boundaries by using the dynamic mode decomposition method, however, there is still a problem of brightness noise. We convert the original video image of the RGB color space to the grayscale and calculate the background feature by using DMD method. When the brightness changes slightly, we find that the background features sometimes changes very violently. As shown in Figure 5 (a), the original video frames brightness changes are not obvious. However, we convert the RGB image to grayscale, and then the amplitude of background mode that calculated by DMD method will change violently due to the brightness changes, which is surely not desired in Figure 5 (b). Therefore, we need to reduce the impact of changes in brightness through the other color space. We choose the HSV color space and we only use the hue channel or saturation channel to calculate the background feature.

In this way, we can reduce the noise of shot detection. In Figure 5 (c), it can be seen that, when we change the color space to HSV, the amplitude of background mode in

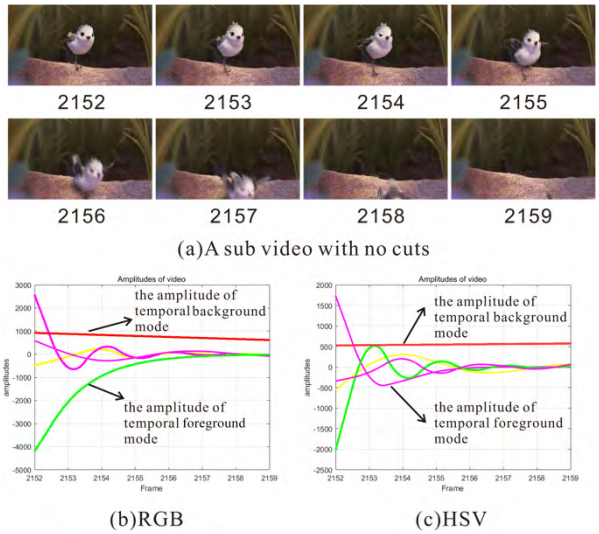


FIGURE 5. The amplitudes by using RGB space (b) and HSV space(c) from video (a).

the video which has no shot boundary is very gentle. And the shot boundary still satisfies the detection condition that the amplitude of background mode changes violently. Using the color space for DMD can eliminate the interference of noise to the shot detection, thus increasing the precision and recall rate.

VII. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we will discuss the detailed implementation and report the algorithm performance. One is the experiment of two difficult problems Q1 and Q2 which Section I refers to. The other is the recall rate and accuracy of the algorithm.

In order to evaluate the performance of shot detection algorithms, our test data contains two different types of video data.

We download some movies and movie trailers. Movie trailers are good source of test data for shot detection, because they often contain a lot of shot transitions in a short sequence, especially the two type of transitions that are difficult to detect: fade and dissolve.

A. APPLICATION IN NON-SHOT BOUNDARIES

In non-shots transitions, the illumination changes sharply, or the foreground objects (or camera) move faster. These factors may cause some algorithms wrongly identify areas as shot boundaries, which are not.

When there is no shot boundary in the group of shots and the brightness changes are more acuity, some algorithms determine this change is the shot boundary. However, we decompose the video by dynamic mode decomposition to obtain temporal background mode and amplitude. Although the intensity of the illumination varies drastically, the background is relatively stable in successive scenes. Obviously, temporal background mode and amplitudes are stable when brightness changes. Figure 6 shows the brightness

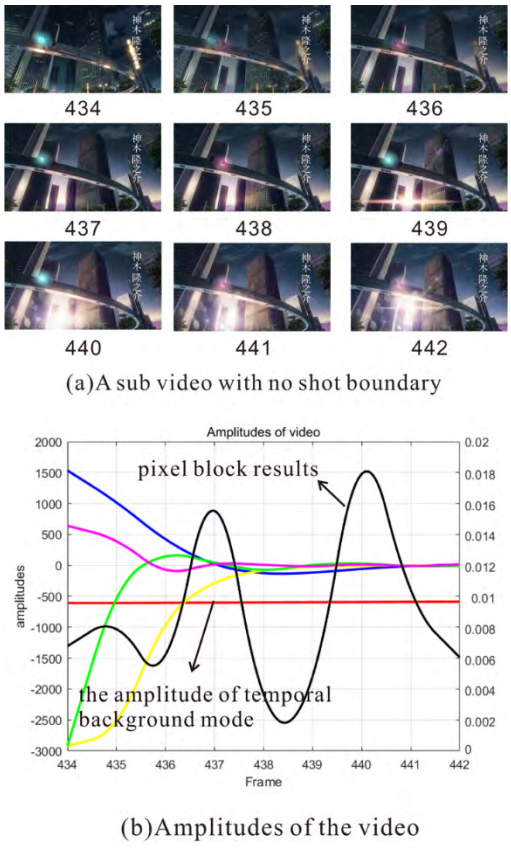


FIGURE 6. The amplitude of temporal background (foreground) mode (b) extracted by DMD from a video without shot boundary, but illumination changes sharply (a).

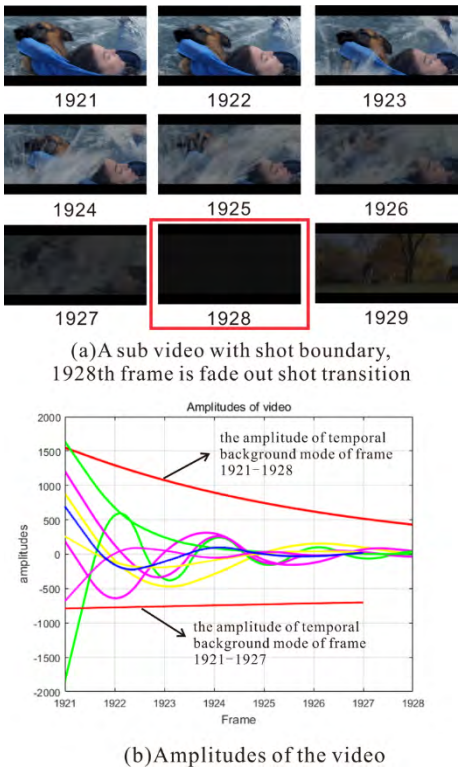


FIGURE 7. The amplitude of temporal background (foreground) mode (b) extracted by DMD from a video without shot boundary, but the foreground object moves too fast (a).

has changed dramatically in nine frames. We found that the pixel block algorithm (black curve in the Figure 6) would determine that 437th and 440th frames are shot boundaries. The 437th and 440th frames have very strong illumination changes, not the shot boundaries. However, using the DMD, we can find that the amplitude of temporal background mode (the red straight line in the Figure 6) is very stable. Therefore, we confirm that there is no shot boundary in this video.

Similarly, when the foreground object (or camera) moves too fast in the sub video, some algorithms determine this change is the shot boundary. Figure 7) is a sub video of the Harry Potter.

In this video, an owl flies swiftly. We found that the pixel block algorithm (black curve in the Figure 7) would determine that 6730th, 6732th and 6734th frames are shot boundaries. However, these three frames are foreground objects moving too fast without the shot boundaries. Therefore, it is very difficult to detect these group of shots. Our algorithm is based on dynamic mode decomposition for temporal background modeling, this problem does not happen. But, the amplitude of temporal background mode (the red straight line in the Figure 7) is terribly stable by using DMD. Therefore, we confirm that there is no shot boundary in this video as well.

B. APPLICATION IN SHOT BOUNDARIES

In shot transitions, the color difference in the different scene is too small like hard cut, the illumination changes slowly like fade, or the foreground objects overlap together like dissolve. Existing algorithms may have missed detection.

Hard cut is the most commonly used in video editing. Hard cut is the two shot directly split joint from head to tail. There are two kinds of hard cuts. In the first one, the luminance differences between two groups of shots is very small, but in the other, the situation is opposite. The traditional method uses the brightness change to detect shot boundary. However, it is impossible to detect the shot boundaries with very small differences in brightness.

The background of the Figure 8 is the beach, and the brightness difference between the shots is not very strong.

Hard cut video as shown in Figure 8. Through the use of DMD shot detection can be seen different amplitude changes in time. Figure 8 (b) the upper red line shows the DMD decomposition and amplitude result of the video between 1777-1784. We can see that the amplitude is very stable and the background is very obvious. However, the video shot has a hard cut in 1785th frame. Using the DMD, we can find that the amplitude has changed dramatically. Therefore, it can be determined that this frame is a hart cut. When there is a significant change in the brightness of the shot transition, the amplitude of the background mode extracted by our method has a drastically changed. Thus, our algorithm can detect two different types of hard cut.

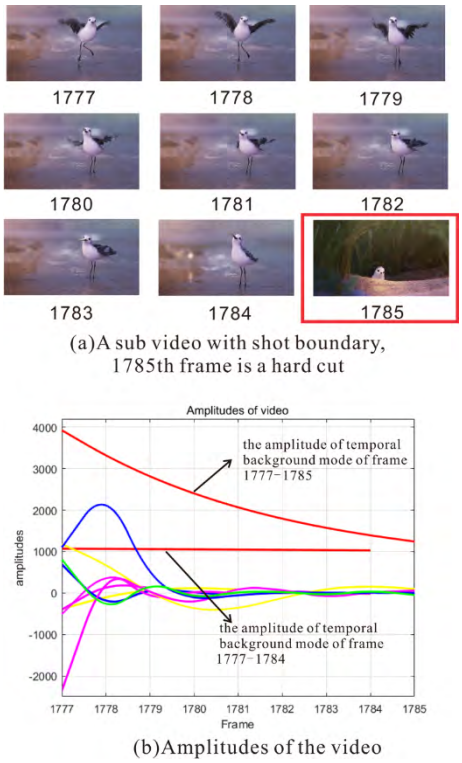


FIGURE 8. (a) A video with a hard cut. (b) The lower red line: from frame 1777-1784 (without shot boundary); the upper red line: from frmae 1777-1785 (with shot boundary).

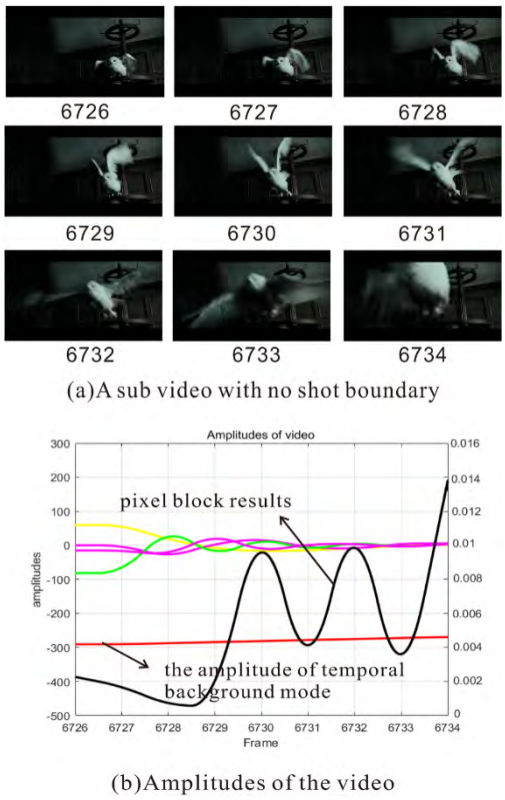


FIGURE 9. The amplitude of temporal background (foreground) mode (b) extracted by DMD from a video with a fade (a).

Fade is another shot transition in video editing. The detection of this shot transition is often very difficult. The color of the shots in fade transition change slowly. Frame rate of the film, TV show or music video is usually 24-30 Fps. In other words, there will be 24-30 frames in the video for one second. In order to create a sense of pause for the viewer, the fade out transition usually lasts for two seconds or more. The traditional method is difficult to detect the shot boundary, because the color change of 35 frames is not very clear.

Using our method can easily detect the fade boundaries. As shown in Figure 9, we get the amplitude of the first four frames of a fade out transition. It represents the background feature of fade and the red line represents background mode. It can be seen that the amplitude is very stable and slow in Figure 9 (b). In the Figure 9, the background of video is the water and dog. In the fade out transition process, the background changes little, but the background disappears at the last frame. When calculating the video from 1921 to 1929 frames, the background feature has a very large amplitude fluctuations. Therefore, it should be a shot transition.

The third method for shot transitions is dissolve. Dissolve, a method to realize shot transition, is a gradual change way from one shot to another. That is the image of the previous shot is superimposed on the image of the next shot in Figure 10 (a). On the boundary of two shots, there will remain the visual features of both the front and the next shot.

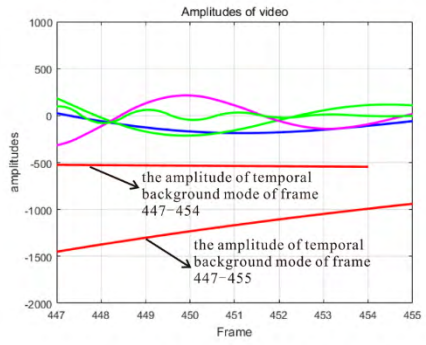
TABLE 1. Our results (A) orinagl data (b) results.

	Name	PIPER	Lala Land	A dog's purpose
(a)	Dur.(sec.)	365	132	128
	Hard cut	38	104	108
	Dissolves	3	8	2
	Fade out	5	1	8
(b)	Correct	40	98	98
	False	5	8	7
	Miss	6	15	20
	Recall	86.96%	86.73%	83.05%
	Precision	88.89%	92.45%	93.33%

Usually, the dissolve boundary will last for 1 to 2 seconds and the length may be 24-48 frames too. Moreover, the length is not always certain, may be very different in accordance with the practice of the director. It can be seen that if the boundary is decided only by the color or the luminance difference, there will be rather a certain amount of errors. Many studies have shown that dissolves cannot be detected easily.

Figure 10 (b) shows that amplitude from frame 447 to frame 454. Although this part of the shot is fade out of transition, it is still not the division of the shots. The amplitude from 447 to 455 frames changes drastically. Thus, it can be clearly detected dissolved shot segmentation.

(a)A sub video with shot boundary, 447th-454th frames are dissolve shot transition



(b)Amplitudes of the video

FIGURE 10. The amplitude of temporal background (foreground) mode (b) extracted by DMD from a video with a dissolve (a).

TABLE 2. Comparison of the proposed method with other algorithms.

Name	Average Recall	Average Precision	F1
Proposed Method	0.856	0.916	0.885
PG Lakshmi Method [47]	0.893	0.863	0.877
E Apostolidis Method [48]	0.885	0.659	0.746
Y. Qi Method [49]	0.940	0.690	0.820
Boccignone Method [50]	0.920	0.890	0.900
J. Yuan Method [22]	0.833	0.865	0.849
color histogram [20]	0.789	0.904	0.842
edge detection [20]	0.702	0.900	0.788
Macroblock [20]	0.753	0.874	0.809
Jordi Mas etc. Method [26]	0.816	0.768	0.792

C. RECALL RATE AND PRECISION

The trial video data set contains 3 video files, and the frame rate of the video is 29 frames per second. The resolution of the video is 640∗480. These videos include hard cuts, dissolves, fades. The total video time is 150 minutes.

The purpose of the shot detection algorithm is to reduce the number of false and missed. Thus, most researchers use precision and recall rates to evaluate the video shot detection algorithm. They are defined as follows:

$$\boldsymbol{Recall} = \frac{\boldsymbol{C}}{\boldsymbol{C}+\boldsymbol{M}} \tag{13}$$

$$\boldsymbol{Precision} = \frac{\boldsymbol{C}}{\boldsymbol{C}+\boldsymbol{F}} \tag{14}$$

where, C is the correct shot boundaries by using our algorithm. M is the missed shot boundaries. F is the false shot boundaries by using our algorithm. The overall performance of the algorithm is judged by $F1$.

$$\boldsymbol{F1} = \frac{\boldsymbol{2 * Recall * Precision}}{\boldsymbol{Recall + Precision}} \tag{15}$$

In addition, we compare the average recall, average accuracy and F1 with other classical shot detection algorithms. The results are shown in Table 1 and 2.

VIII. CONCLUSIONS AND FUTURE WORK

A dynamic mode decomposition based video shot detection is proposed in this paper which shows good performance on the different shot transitions. Through the modes extracted by DMD, we can get the amplitude of temporal background mode of the video so as to know whether the content in the video is related. If the content belongs to different shots, there will be a suddenly change of the amplitude of temporal background mode so that we can get the shot boundaries. Meanwhile, our method can effectively reduce the rate of error shot boundary detection for the illumination changes sharply, or the foreground object (or camera) moves faster. It is a universal method for shot transitions, like hard cuts, fades and dissolves. We have obtained good results for applying DMD on the shot detection problem. This result also can give a reference to computer vision technology used in the intelligent transportation system, intelligent vehicle, and so on [41]–[46].

Finally, the experimental data of actual processing are analyzed, and some issues are discussed, forming the objects for future work. We will work on the automatically adaptive threshold which is selected to evaluate and decide where the shot boundary is. If the threshold can be set through the computing of the computer without mutual analysis, the shot detection will go ahead a little further.

REFERENCES

[1] S. Tippaya, S. Sitjongsataporn, T. Tan, M. Khan, and K. Chamnongthai, "Multi-modal visual features-based video shot boundary detection," *IEEE Access*, vol. 5, pp. 12563–12575, 2017.

[2] Q. Li, F. He, T. Wang, L. Zhou, and S. Xi, "Human pose estimation by exploiting spatial and temporal constraints in body-part configurations," *IEEE Access*, vol. 5, pp. 443–454, 2017.

[3] Z. Zhang, T. Jing, J. Han, Y. Xu, and X. Li, "Flow-process foreground region of interest detection method for video codecs," *IEEE Access*, vol. 5, pp. 16263–16276, 2017.

[4] Y. Li, R. Xia, Q. Huang, W. Xie, and X. Li, "Survey of spatio-temporal interest point detection algorithms in video," *IEEE Access*, vol. 5, no. 2, pp. 10323–10331, Feb. 2017.

[5] M. S. Hossain and G. Muhammad, "An emotion recognition system for mobile applications," *IEEE Access*, vol. 5, pp. 2281–2287, 2017.

[6] P. Botsinis, Y. Huo, D. Alanis, Z. Babar, S. Ng, and L. Hanzo, "Quantum search-aided multi-user detection of IDMA-assisted multi-layered video streaming," *IEEE Access*, vol. 5, pp. 23233–23255, 2017.

[7] S. Choudhury, P. K. Sa, S. Bakshi, and B. Majhi, "An evaluation of background subtraction for object detection vis-a-vis mitigating challenging scenarios," *IEEE Access*, vol. 4, pp. 6133–6150, 2017.

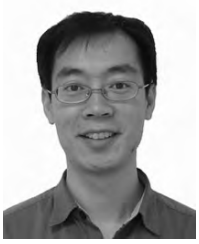
[8] S. Li, Z. Qin, and H. Song, "A temporal-spatial method for group detection, locating and tracking," *IEEE Access*, vol. 4, pp. 4484–4494, 2016.

[9] A. Smeaton, P. Over, and A. Doherty, "Video shot boundary detection: Seven years of TRECVid activity," *Comput. Vis. Image Understand.*, vol. 114, no. 4, pp. 411–418, 2010.
[10] Y. N. Li, Z. M. Lu, and X. M. Niu, "Fast video shot boundary detection framework employing pre-processing techniques," *IET Image Process.*, vol. 3, no. 3, pp. 121–134, Jun. 2009.
[11] H. Wang and S. K. Nguang, "Multi-target video tracking based on improved data association and mixed Kalman/H_∞ filtering," *IEEE Sensors J.*, vol. 16, no. 21, pp. 7693–7704, Nov. 2016.
[12] H. Feng, W. Fang, S. Liu, and Y. Fang, "A new general framework for shot boundary detection and key-frame extraction," *J. Tsinghua Univ.*, vol. 8, no. 2, pp. 121–126, 2005.
[13] O. A. Olumodeji, A. P. Bramanti, and M. Gottardi, "A memristive pixel architecture for real-time tracking," *IEEE Sensors J.*, vol. 16, no. 22, pp. 7911–7918, Nov. 2016.
[14] J. Ren, J. Jiang, and J. Chen, "Shot boundary detection in MPEG videos using local and global indicators," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 8, pp. 1234–1238, Aug. 2009.
[15] H. Fang, J. Jiang, and Y. Feng, "A fuzzy logic approach for detection of video shot boundaries," *Pattern Recognit.*, vol. 39, no. 11, pp. 2092–2100, Nov. 2006.
[16] A. Amiri and M. Fathy, "Video shot boundary detection using decomposition and Gaussian transition detection," *EURASIP J. Adv. Signal Process.*, vol. 2009, p. 509438, Dec. 2010. [Online]. Available: https://link.springer.com/article/10.1155/2009/509438
[17] R. Hannane, A. Elboushaki, K. Afdel, P. Naghabhushan, and M. Javed, "An efficient method for video shot boundary detection and keyframe extraction using SIFT-point distribution histogram," *Int. J. Multimedia Inf. Retr.*, vol. 5, no. 2, pp. 89–104, 2016.
[18] B. Han, Y. Hu, G. Wang, W. Wu, and T. Yoshigahara, "Enhanced sports video shot boundary detection based on middle level features and a unified model," *IEEE Trans. Consum. Electron.*, vol. 53, no. 3, pp. 1168–1176, Aug. 2007.
[19] P. J. Schmid, "Application of the dynamic mode decomposition to experimental data," *Experim. Fluids*, vol. 50, no. 4, pp. 1123–1130, 2011.
[20] P. Browne, A. Smeaton, N. Murphy, N. O'Connor, S. Marlow, and C. Berrut, "Evaluating and combining digital video shot boundary detection algorithms," *DORAS*, vol. 243, no. 2000, pp. 1–8, 1999.
[21] J. S. Boreczky and L. A. Rowe, "Comparison of video shot boundary detection techniques," *J. Electron. Imag.*, vol. 5, no. 2, pp. 122–128, Apr. 1996.
[22] J. Yuan *et al.*, "A formal study of shot boundary detection," *IEEE Trans. Circuits Syst. for Video Technol.*, vol. 17, no. 2, pp. 168–186, Feb. 2007.
[23] I. Koprinska and S. Carrato, "Temporal video segmentation: A survey," *Signal Process., Image Commun.*, vol. 16, no. 5, pp. 477–500, 2001.
[24] J. Bescos, G. Cisneros, J. M. Martinez, J. M. Menendez, and J. Cabrera, "A unified model for techniques on video-shot transition detection," *IEEE Trans. Multimedia*, vol. 7, no. 2, pp. 293–307, Apr. 2005.
[25] S. Manjunath, D. S. Guru, M. G. Suraj, and B. S. Harish, "A non-parametric shot boundary detection: An Eigen gap based approach," *J. Cell Sci.*, vol. 118, no. 9, pp. 1923–1934, 2011.
[26] J. Mas and G. Fernandez, "Video shot boundary detection based on color histogram," Digital Television Center (CeTVD), La Salle School Eng., Ramon Llull Univ., Barcelona, Spain, Notebook Papers TRECVID2003, 2003.
[27] F. Idris and S. Panchanathan, "Review of image and video indexing techniques," *J. Vis. Commun. Image Represent.*, vol. 8, no. 2, pp. 146–166, 1997.
[28] S. Natarajan, "An efficient video segmentation algorithm with real time adaptive threshold technique," *Int. J. Signal Process.*, vol. 2, no. 4, pp. 13–28, 2009.
[29] B. V. Patel and B. B. Meshram, "Content based video retrieval systems," *Int. J. Ubicomp*, vol. 3, no. 2, p. 13, 2012.
[30] H. Wang, A. Divakaran, A. Vetro, S. Chang, and H. Sun, "Survey of compressed-domain features used in audio-visual indexing and analysis," *J. Vis. Commun. Image Represent.*, vol. 14, no. 2, pp. 150–183, 2013.
[31] S. K. Biswas and P. Milanfar, "One shot detection with laplacian object and fast matrix cosine similarity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 3, pp. 546–562, Mar. 2016.
[32] B. S. Rashmi and H. S. Nagendraswamy, "Abrupt shot detection in video using weighted edge information," in *Proc. Int. Conf. Informat. Analytics*, 2016, pp. 1–5.
[33] H. Lin, J. Deng, and B. Woodford, "Shot boundary detection using multi-instance incremental and decremental one-class support vector machine," in *Proc. Pacific–Asia Conf. Knowl. Discovery Data Mining*, 2016, pp. 165–176.
[34] F. Desobry, M. Davy, and C. Doncarli, "An online kernel change detection algorithm," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2961–2974, Aug. 2005.
[35] J. Cao and A. Cai, "A robust shot transition detection method based on support vector machine in compressed domain," *Pattern Recognit. Lett.*, vol. 28, no. 12, pp. 1534–1540, 2017.
[36] P. J. Schmid, "Dynamic mode decomposition of numerical and experimental data," *J. Fluid Mech.*, vol. 656, no. 10, pp. 5–28, 2010.
[37] C. W. Rowley, I. Mezic, S. Bagheri, P. Schlatter, and D. S. Henningson, "Spectral analysis of nonlinear flows," *J. Fluid Mech.*, vol. 641, pp. 115–127, Dec. 2009.
[38] P. J. Schmid, L. Li, M. P. Juniper, and O. Pust, "Applications of the dynamic mode decomposition," *Theor. Comput. Fluid Dyn.*, vol. 25, nos. 1–4, pp. 249–259, 2011.
[39] J. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, "On dynamic mode decomposition: Theory and applications," *J. Comput. Dyn.*, vol. 1, no. 2, pp. 391–421, 2015.
[40] C. Bi, Y. Yuan, R. Zhang, Y. Xiang, Y. Wang, and J. Zhang, "A dynamic mode decomposition based edge detection method for art images," *IEEE Photon. J.*, vol. 9, no. 6, pp. 1–13, Jun. 2017.
[41] R. Zhang, Y. Ma, F. You, T. Peng, Z. He, and K. Li, "Exploring to direct the reaction pathway for hydrogenation of levulinic acid into gvalerolactone for future clean-energy vehicles over a magnetic Cu-Ni catalyst," *Int. J. Hydrogen Energy*, vol. 42, no. 40, pp. 25185–25194, 2017.
[42] R. Zhang, Z. He, H. Wang, F. You, and K. Li, "Study on self-tuning tyre friction control for developing main-servo loop integrated chassis control system," *IEEE Access*, vol. 5, pp. 6649–6660, 2017.
[43] R. Zhang, J. Wu, L. Huang, and F. You, "Study of bicycle movements in conflicts at mixed traffic unsignalized intersections," *IEEE Access*, vol. 5, pp. 10108–10117, 2017.
[44] L. Yang, B. Wang, R. Zhang, H. Zhou, and R. Wang, "Analysis on location accuracy for the binocular stereo vision system," *IEEE Photon. J.*, vol. 10, no. 1, pp. 1–16, Jan. 2018.
[45] H. Huang, A. Yang, L. Feng, G. Ni, and P. Guo, "Indoor positioning method based on metameric white light sources and subpixels on a color image sensor," *IEEE Photon. J.*, vol. 8, no. 6, Dec. 2016, Art. no. 6806110.
[46] H. Maestre, A. J. Torregrosa, and J. Capmany, "IR Image Upconversion under Dual-Wavelength Laser Illumination," *inIEEE Photon. J.*, vol. 8, no. 6, pp. 1–8, 2016.
[47] G. Lakshmi and S. Domnic, "Walshhadamard transform kernel-based feature vector for shot boundary detection," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5187–5197, Dec. 2014.
[48] E. Apostolidis and V. Mezaris, "Fast shot segmentation combining global and local visual descriptors," in *Proc. IEEE Int. Conf. Acoust.*, May 2014, pp. 6583–6587.
[49] Y. Qi, A. Hauptmann, and T. Liu, "Supervised classification for video shot segmentation," in *Proc. IEEE Conf. Multimedia Expo*, vol. 2. Jul. 2003, pp. 689–692.
[50] G. Boccignone, A. Chianese, V. Moscato, and A. Picariello, "Foveated Shot Detection for Video Segmentation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 3, pp. 336–337, Mar. 2005.

CHONGKE BI received the B.Sc. (Eng.) and M.Sc. (Eng.) degrees from Shandong University, China, in 2004 and 2007, respectively, and the Ph.D. (Sci.) degree from the University of Tokyo, Japan, in 2012. He was a Researcher with RIKEN, Japan, where he was focused on the research in the field of visual analysis of big data on supercomputer from 2012 to 2016. He is currently with Tianjin University. His current research interests include image processing, big data, visualization, and computer graphics.

YE YUAN received the bachelor's and master's degrees from the School of Computer Software, Tianjin University, in 2012 and 2014, respectively, where he is currently pursuing the Ph. D. degree. His research interests include video analysis and multimedia processing.

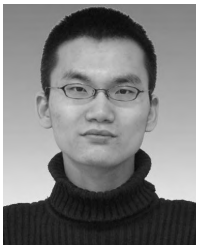
JIAWAN ZHANG received the master's and Ph.D. degrees in computer science from Tianjin University in 2001 and 2004, respectively. He is currently a Professor with the School of Computer Software and an Adjunct Professor with the School of Computer Science and Technology, Tianjin University. He has published over 50 academic papers in peer-reviewed journals and conferences. He holds five patents and five software copyrights. His main research interests include computer graphics and realistic image synthesis. He received the IBM Global Faculty Award for his contribution to the software industry in 2005 and 2006. He was selected as the top 10 Distinguished Younger of Tianjin University in 2007 and as the Best Young Teacher of Tianjin University in 2008.



YUN SHI received the B.S. and Ph.D. degrees in civil engineering from the University of Tokyo, Japan, in 2004 and 2007, respectively. He is currently a Professor with the Key Laboratory of Resources Remote Sensing and Digital Agriculture, Ministry of Agriculture/Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing, China. His current research interests include remote sensing, smart agriculture, pattern recognition, urban 3-D modeling, and multiple sensor data fusion.

YIQING XIANG is currently pursuing the bachelor's degree with the School of Computer Software, Tianjin University. Her research interests include graphic processing and visual analysis.

YUEHUAN WANG is currently pursuing the bachelor's degree with the School of Computer Software, Tianjin University. Her research interests include image processing and machine learning.



RONGHUI ZHANG received the B.Sc. (Eng.) degree from the Department of Automation Science and Electrical Engineering, Hebei University, Baoding, China, in 2003, the M.S. degree in vehicle application engineering from Jilin University, Changchun, China, in 2006, and the Ph.D. (Eng.) degree in mechanical and electrical engineering from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, in 2009. In 2011, he was a Post-Doctoral Researcher with INRIA, Paris, France. He is currently a Research Fellow with the Guangdong Key Laboratory of Intelligent Transportation System and Research Center of Intelligent Transportation Systems, School of Engineering, Sun Yat-sen University, Guangzhou, Guangdong, China. His current research interests include computer vision, intelligent control, and ITS.