An Illumination Invariant Measure in Detecting Scene Breaks*

LI Da-long, LU Han-qing

(*National Laboratory of Pattern Recognition, Institute of Automation, The Chinese Academy of Sciences, Beijing* 100080, *China*)

E-mail: dlli@nlpr.ia.ac.cn; luhq@nlpr.ia.ac.cn

http://www.ia.ac.cn

Received July 19, 2000; accepted February 20, 2001

Abstract: Illumination variation inherent in the video production process is a serious problem in video shot detection. It disturbs the precision of many shot detection algorithms. The computing cost of the current illumination invariant measures is high thus lower the speed of the algorithm. A new measure is proposed in the paper to detect shots in video. The new measure is based on the edges in the frames. The edge maps are extracted first then it is thresholded and the authors take the area of the background as the illumination invariant measure to perform the detection of shots. The effectiveness of the measure had been validated by theoretic analysis and experimental results. Compared with the existing measures, the new measure improves the speed and reserves the precision of the detecting algorithms.

Key words: video partitioning; histogram; illumination variation; dissolve; wipe

With the dramatical increase of video data volume, it has become important to archive and access the video data in several important application fields such as VOD (video on demand), DLI (digital library), etc. Video is hard to efficiently browse and retrieve, because it combines all the other media information into a single bit stream. Computer vision technologies promise to allow content-based browsing of image sequences. For example, we may be able to jump from one shot to another if the first one does not interest us. This will require the algorithms to automatically detect the boundaries between shots. The boundaries include cuts, fades, dissolves and wipes. Fades could be taken as special case of dissolves.

The boundaries (scene breaks) mark the transitions from one sequence of consecutive images to another. According to the duration of shot boundaries, there are two types: camera breaks and gradual transitions. Camera break (cut) is an instantaneous transition from one scene to the next. Basically there are two kind of gradual transition: wipe and dissolve. A wipe is a moving boundary line crossing the screen such that one shot gradually replaces another; (there is a wipe in the Fig. 4); a dissolve superimposes two shots where one shot gradually lightens while the other fades out slowly. An example of dissolve can be found in Fig. 5. Fade is special case of dissolve. A

* Supported by Foundation of State Key Basic Research 973 Program under Grant No. G1998030500 (973 国家重点基础研究发展项目基金); the National High Technology Development Program of China under Grant No. 863-306-QN99-2 (国家 863 高科技发展计划)

LI Da-long was born in 1976. He is a M. S. student at the National Laboratory of Pattern Recognition, Institute of Automation, The Chinese Academy of Sciences. His current research interests are video/image processing, computer vision, pattern recognition as well as DSP. **LU Han-qing** was born in 1961. He is a professor and doctoral supervisior at the National Laboratory of Pattern Recognition, Institute of Automation, The Chinese Academy of Sciences. His current research interests are image analysis and understanding, pattern recognition, multimedia, medical image and so on.

fade is a gradual transition between a scene and a constant image (fade in) or between a constant image and a scene (fade in). Linear model is reasonable in describing dissolves and fades. Shot detection is a critical step in content-based video browse and retrieval. Recently, substantial efforts have been devoted to shot-based video analysis. A survey on this topic can be found in Ref. [1].

An assumption often made is that the content should remain nearly the same from one frame to the next within one camera shot. So, In general, shot boundaries can be detected by employing a difference metric to measure the change between two consecutive frames. A shot boundary is declared if the difference exceeds a certain threshold. Pixel- or Block- based temporal image difference[2,3], or difference of gray and color histograms[4,5] was supposed to be the measure metric. Histograms are robust to object motion. And they are easy to compute. So they have been widely used in shot-based video analysis. And several researchers claim that this measure can achieve good trade-off between accuracy and speed. Unfortunately there is a problem of segmenting the film into a sequence of shots based on difference of histograms when illumination varies.

The Color Ratio Histogram is adopted by W.X. Kong[6] to act as an illumination in variant metric to solve the problem, but it is expensive in computing. The success of Color Ratio Histogram in illumination variation is based on the fact that when a pixel is lightened, the pixels around that one are also lightened, but the ratio changes little. Similarly, the proposed metric in the paper also computes the difference of the gray level of the pixels, but only the simplest neighborhood is used to extract the edge map of the image. Then the area of the background in the edge map is computed and the difference of the area of the consecutive frames is used to measure the similarity of the content. Compared to the calculation of the histogram difference of two consecutive, it is more economic regarding the computing cost. Area is simply a scalar quantity while both color/gray histogram and ratio histogram are vectors.

Ramin Zabih[7] proposed a feature-based algorithm for detecting and classifying scene breaks. His algorithm is also robust when illumination varies since they detect the appearance of intensity edges that are distant from the edges in the previous frame. The edges are invariant when lighting conditions change. A global motion computation is used to handle camera or object motion. The algorithm runs slow. It is reported in the paper that an initial implementation runs at approximately 2 frames per second on a sun workstation.

The paper is organized as follows. In Section 1, we begin with a survey of some related work especially those of Kong and Ramin Zabih. Kong's suggestion works well however the speed is a little low. Actually we need not use a metric as precise as color ratio histogram. Low speed is the problem of Ramin Zabih's method too. In Section 2, the ABEM is introduced to address the illumination problem and its robustness is illustrated. The effectiveness of the methods is validated by experiments on some real-world video sequences. Some experimental results are also discussed in Section 3. Concluding remark is made in Section 4 that addresses the extension of the method as well as the limitation of it.

1 Related Work

Many methods have been proposed to deal with video segmentation to make it easy to browse and retrieve video. Besides twin comparison[4], there is STDD (single threshold double directions) and Multi-scale hierarchy video segmentation that can be found in some of my papers[8,9]. These methods are effective in detecting gradual transitions. Due to the desirable quality of robustness to object motion, the gray or color histograms are widely used in video segmentation. Unfortunately they ignore the problem of illumination variation inherent in the video production process. So they often make wrong judgment when the incident illumination varies. The illumination variation greatly disturbs the detections. Even simple lighting changes will result in abrupt changes in histograms. This limitation might be overcome by preprocessing with a color constancy algorithm. In Ref. [10], Wei Jie *et al.*

proposed a color-channel-normalization method. Then they reduced the three-dimension color to two-dimension chromatic and defined a two-dimension chromatic histogram. Their method can discount simple spectral changes of illumination. But it is computationally expensive and cannot do with spatial changes of illumination.

In Ref. [11], Funt *et al.* studied the use of color ratio histograms as features for indexing image database. And it was adopted in shot detection. Color ratio histogram can be formulated as following:

$$H(i,j,k)=\sum_{x,y} z(x,y),$$

$$z(x,y)=\begin{cases}1, & \text{if } d_R(x,y)=i\ d_G(x,y)=j\ d_B(x,y)=k\\ 0, & \text{else}\end{cases}$$

$$d_k(x,y)=\nabla^2 i_k(x,y),k=R,G,B$$

$$i_k(x,y)=\log f_k(x,y),k=R,G,B$$

where $f(x,y)$ is the RGB color value at position (x,y), $i(x,y)$ is their logarithm and $d(x,y)$ is the Laplacian difference of $i(x,y)$. Color ratio histogram is the histogram of $d(x,y)$. Unfortunately, the computing cost is too high. RGB varies in the range {0 255} because we need to call the function of log( ). So it is necessary to add 1 to every value. Therefore we get {1 256}. The max possible value for d is then 4 * LOG(256). The integral part of it is 22, the min is −22. Suppose the scale of an image is 288 * 352 pixels. Then 45 * 45 * 45 * 286 * 350 (9121612500) loops are needed to get the color ratio histogram for that image, but there are many frames even in a very short video. It is easy to notice that the computing time is too much.

Let us look back at the original of the color ratio histogram. It is proposed in database indexing therefore we need to guarantee the precision. It is worth of making such computation. However in shot detection, such precision is not necessary. During a shot, the contents of the frames are very similar to each other. Such similarity can be measured by accurately or inaccurately. Whatever the measurements are, the results are the same. However the computing costs of different measurements are not same. Accurate measurements need a lot of computation thus lower the speed of the algorithm. But inaccurate measurements, are more economic in computing. This is the motivation of the proposal of the new measurements.

Ramin Zabih's feature-based algorithm works well even there exist illumination variations. In his algorithm, edge dilation follows edge detections. His algorithm can classify different wipes. However the important thing for us is to get the shots. Dissolves and wipes are just the transitions of shots. They make no sense in video content understanding. Suppose we exchange a wipe with a dissolve, can the meaning of the video be effected? Of course not.

2 Abem: An Illumination Invariant Metric

Realizing that the edge map remains unchanged when lighting condition varies, ABEM (Area of the Boundary in Edge Map) is proposed to detect shots. ABEM is worked out in the following way: the color image is firstly turned into a gray-scale image, then the object edges are extracted by differentiating the pixels in the image line by line and then thresholding it. A binary image (edge map) can be obtained. Finally, count the pixels contributing to the edge in the edge map and we can get ABEM. Video segmentation is based on the comparison of ABEM of the consecutive images. The procedure to calculate ABEM can be formulated as follows:

$$ABEM=\sum_{x,y} b(x,y),$$

$$b(x,y)=\begin{cases}1, & \text{if } z(x,y)=0\\ 0, & \text{else}\end{cases}$$

$$z(x,y)=\begin{cases}255, & \text{if } d(x,y)>T\\ 0, & \text{else}\end{cases}$$

$$d(x,y)=\nabla i(x,y)$$

where $i(x,y)$ is the Gray Level at position (x,y), and $d(x,y)$ is the difference of $i(x,y)$. Z makes up the bilevel image of the object edge.

Table 1 shows the ABEMs of some typical images taken from a sequence. Frame100 and Frame 101 are two frames inside a shot, Frame 189 and Frame 190 are two inside a shot, but there is an illumination variation between them. Then they are compared in Table 2. It is easy to perceive that inside a shot ABEM is stable, and when illumination varies, it is still insensitive. But when a camera break happens, it does change considerably, thus the efficiency of the video segmentation is improved. The illumination variation is illustrated in Fig. 1.

Table 1 ABEM of some typical images

Frame No.	ABEM
100	99477
101	99424
189	100593
190	100422

ABEM: area of boundary in edge map

Table 2 Comparison between ABEM and GHD

Cases	Frames	DABEM	DGH
Inside a Shot	100,101	53	6410
Camera Break	100,189	1116	49982
Illumination Variation	189,190	171	41392

DABEM: difference of ABEM

DGH: difference of gray level histogram

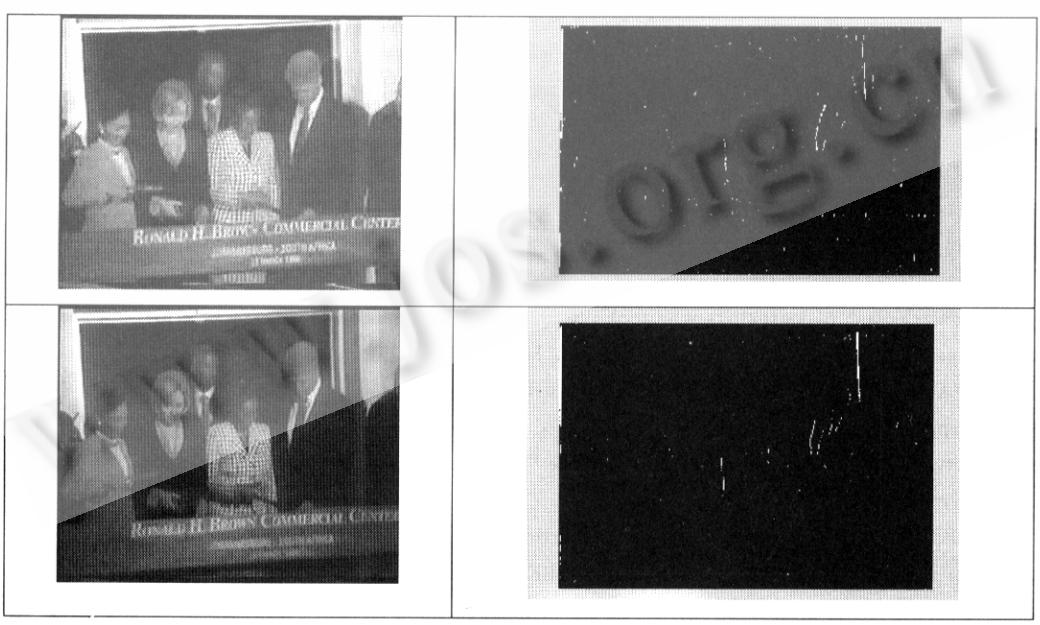

Fig. 1 An example of the illumination variation

As we can see that the edge maps are nearly the same though one of the frames is obviously brighter than the other. From the table we can still notice that false alarm will be made based on gray level histogram comparison. In order to evaluate the efficiency of the metric, we further define R_1 and R_2. R_1 is the ratio between the difference of two frames belonging to different shots and the difference two frames inside a shot; R_2 is the ratio of the difference of two frames belonging to different shots and the difference of two frames inside a shot with an illumination variation. Clearly, the larger R_1 and R_2 are, the better. Because it becomes rather easy to distinguish them thus to segment the video efficiently when the R_1 and R_2 are larger. Under the metric of AEBM, they are both larger than under the metric of gray level histogram as shown in Table 3.

Table 3 Metric evaluation

	$R1$	$R2$
ABEM	21.056	6.526
GH	7.798	1.208

$R1$: D_k(camera break)/D_k(inside a shot) k=ABEM, GH

$R2$: D_k(camera break)/D_k(illumination variation) k=ABEM,GH

The comparison of the computation cost between AEBM and CRH is further compared in Table 4. To get the AEBM for an image, the pixel in the image is visited only once, but CRH need 45 * 45 * 45 (91125) times, so the speed is improved greatly.

Table 4 Computation cost contrast between AEBM and CRH

	AEBM	CRH
Loop time of differentiation	$r\times(c-1)$	$3^{(r-1)\times(c-1)}$
Loop time of counting	1	45 * 45 * 45
Is Log( ) called?	no	Yes

Suppose the image size is $r\times c$

AEBM: area of boundary in edge map

CRH: color ratio histogram

3 Expermental Results

To test the effectiveness of the proposed metric, experiments are conducted on some video sequences that contain illumination variations as well as camera breaks and gradual transitions.

The approach has been validated by the designed test. Some results are illustrated in Figs. 2 and 3. The test consequence is a segment from a news report. In the sequence there are many illumination variations caused by the photo-chemistry-flash of journalists who took photograph in a ceremony. They are all discarded by employing the proposed metric. Therefore no false alarm is made. Figure 2 gives the curve of the ABEM difference of the consecutive frames. Based on the figure, three camera breaks in the segmentation of the video are found. The first one is a scene change, the content of the picture is changed from the reporter to the event that the reporter is reporting. The other two are due to the appearance of the title "Johannesburg, South Africa" and the disappearance of the title as shown in Fig. 3.

To validate the effectiveness of ABEM in detecting gradual transitions, both wipes and dissolves are tested using ABEM. From Figs. 4 and 5, we can see that ABEM is sensitive to both wipes and dissolves. Therefore ABEM is able to be used to detect gradual transitions successfully.

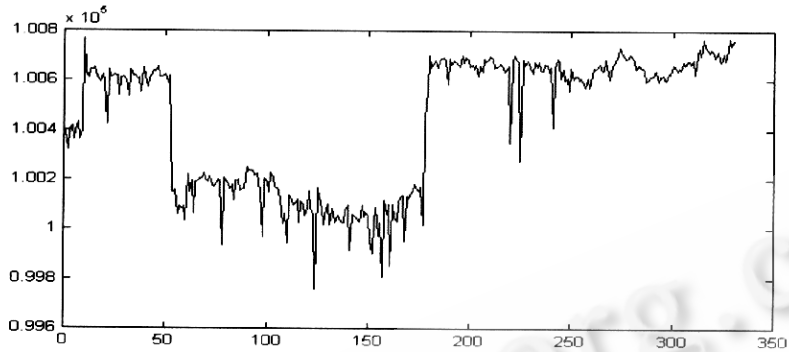

Fig. 2 The ABEM of a sequence from a news report

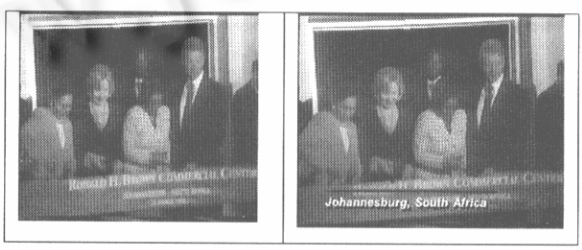

Fig. 3 The apperance of the title

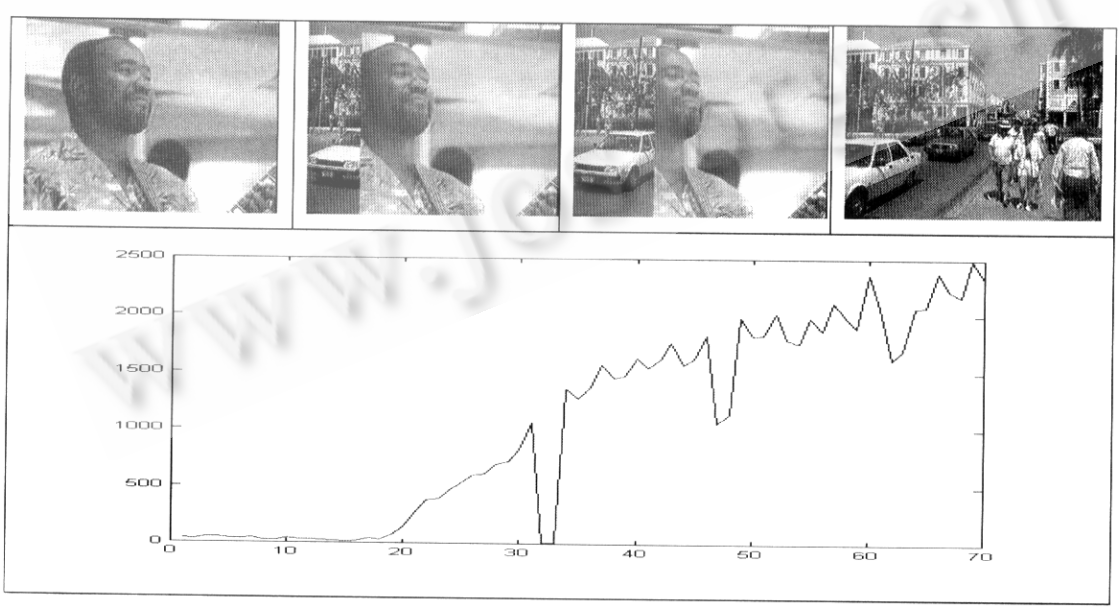

Fig. 4 ABEM is sensitive to wipe

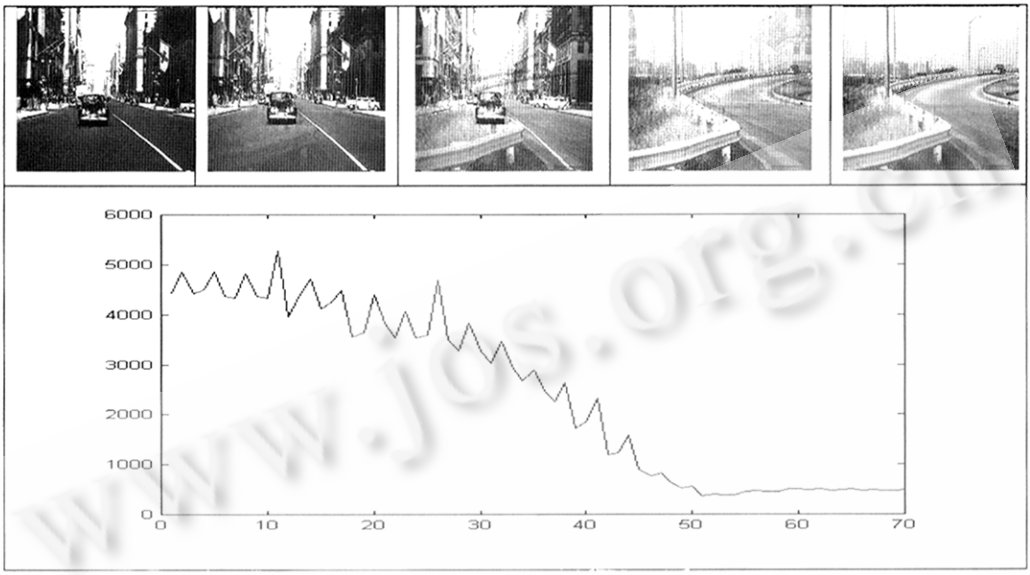

Fig. 5 ABEM is sensitive to dissolve

4 Conclusion Remark

In this paper, a new illumination invariant metric named ABEM is proposed to measure the similarity of the content of images and its advantages are validated. ABEM is insensitive to lighting condition variation but it can reflect the shot shifts no matter whether it is a camera break or gradual transition (wipe and dissolve). ABEM is economic in computation compared with color ratio histogram.

The limitation of the measures is that false alarms due to camera motions are still inevitable. ABEM can only deal with illumination variations. It is beyond its ability when large camera motions or object motions occur. In fact, other similar methods such as color ratio histogram are also useless under such situation. To meet the problem, we need other approaches to get the job done. One of solutions is proposed in my paper[12]. In that paper, we discard large object/camera motions by picking up wipes/dissolves. Both wipes and dissolves are regular transitions. However camera/object motions are random.

References:

[1] Boreczky, J. S., Rowe, L. A. Comparison of video shot boundary detection techniques. In: Sethi, I. K., Jain, R. C., eds. Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases IV. Bellingham: SPIE Press, 1996. 170~179.

[2] Otsuji, K., Tonomura, Y., Ohba, Y. Video browsing using brightness data. In: Tzou, K. H., Koga, T., eds. Proceedings of the SPIE on Visual Communications and Image Processing. Bellingham: SPIE Press, 1991. 980~989.

[3] Nagasaka, Tanaka, Y. Automatic video indexing and full-video search for object appearances. Proceedings of the 2nd Working Conference on Visual Database Systems. 1991. 119~133.

[4] Zhang, H., Kankanhalli, A., Smoliar, S. Automatic partitioning of full-motion video. Multimedia Systems, 1993,1:10~28.

[5] Sethi, I. K., Patel, N. A statistical approach to scene change detection. In: Niblack, W., Jain, R. C., eds. Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases. Bellingham: SPIE Press, 1995. 329~338.

[6] Kong, W. X., Ding, X. F., Lu, H. Q., *et al*. Improvement of shot detection using illumination invariant metric and dynamic threshold selection. In: Huijsmans, D. P., Smeulders, A. W. M., Lecture Notes in Computer Science 1614. Springer-Verlag Berlin, Heidelbegy, 1999. 277~282.

[7] Zabih R., Miller J., Mai, K. A feature-based algorithm for detecting and classifying production effects. ACM Journal of Multimedia Systems, 1999,7(2):119~128.

[8] LI, Da-long, LU, H. Q. Video segmentation through STDD (Single Threshold Double Directions). In: Hamza, M. H., ed. The IASTED Proceedings of the International Conference on Modeling and Simulation (MS2000). Calgary, ACTA Press, 2000. 372~375.

[9] LI, Da-long, LU, H. Q. Multi-Scale hierarchy video segmentation. Proceedings of the 1st IEEE EIT Conference. Chicago, USA, 2000.

[10] Wei, J., Drew, M. S., LI, Z.-N. Illumination-invariant video segmentation by hierarchical robust thresholding. In: Sethi, I. K., Jain, R. C., eds. Proceedings of the SPIE Conference in Storage and Retrieval for Image and Video Databases. Bellingham: SPIE Press, 1998. 188~201.

[11] Funt, B. V., Finlayson, G. D. Color constant color indexing. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995,17:522~529.

[12] LI, Da-long, LU, Han-qing. Model based video segmentation. Proceedings of the 2000 IEEE Workshop on Signal Processing Systems. Lafayette, USA: IEEE Press, 2000.

一种用于镜头检测的光照不变测度

李大龙, 卢汉清

(中国科学院 自动化研究所 模式识别国家重点实验室,北京 100080)

摘要: 光照问题是镜头检测中的一个重要的问题.光照变化严重干扰了许多镜头检测的精度.而现有的光照不变的帧间相似性测度计算代价高而降低了算法的速度.为此,提出了一种新的光照不变测度.并将它用于视频镜头检测.这种新的测度是基于图像中的边缘信息.它先提取了图像中的边缘,然后二值化,最后把背景的面积作为光照不变测度.理论分析及实验数据都证明了该测度的有效性.与同类测度相比,提高了速度而没有牺牲精度.

关键词: 视频分割;直方图;光照变化;溶解;扫换

中图法分类号: TP391 **文献标识码**: A