

A Self-Adaptive Shot Boundary Detection Method Based on Neighbor Frame Difference and Non-neighbor Frame Difference

Junxiao Liu

Library, Shandong Normal University, Jinan, China
liujx81@163.com

Abstract

According to the characters of abrupt change and gradual change, we put forward a self-adaptive method of shot boundary detection, which is based on neighbor frame difference and non-neighbor frame difference. By computing neighbor frame difference, this method can effectively detect abrupt changes. By computing non-neighbor frame difference, gradual changes can be effectively detected. Integrating the two methods, it can almost detect all shot boundaries no matter what the shot change type is. Using this method, we does not need to set global threshold and has high adaptability with low computation cost. Experimental results show it is adaptive and efficient.

Keywords: *neighbor frame difference, non-neighbor frame difference, self-adaptive method, shot boundary detection*

1. Introduction

With rapid development of network and computer technology, the dull text information cannot satisfy people's demands any more. Multimedia contents have gradually entered people's life, and digital video has been applied more and more extensively. Therefore, there are more and more relevant researches on video processing. Shot detection is the first step of video processing and it is the foundation for the follow-up content analysis, classification, indexing and query at high levels. Accuracy of shot detection will directly influence effects of follow-up processing. Thereby, study on shot boundary detection has important practical significance.

Shot boundary change modes include abrupt change and gradual change. By directing at characteristics of abrupt change and gradual change, many domestic and overseas researchers have proposed different shot boundary detection algorithms. Global single threshold value method [1], second order difference method [2], sliding window method, and second order frame difference method [3] have gained a good effect in abrupt change shot, but its detection effect for gradual change shot is not that ideal. In order to improve detection effect for gradual change shot, literature [4] gives a digital video cheat shot detection algorithm; literature [5-6] puts forward that video editing model can be used to detect fade-in, fade-out and lap dissolve effects. These methods are only effective for some special gradual change effects. In order to increase the comprehensive detection effect of abrupt change shot and gradual change shot, literature [7] provides a double threshold method which integrates abrupt change and gradual change into one method for realization. However, it is difficult to gain a good detection effect for shot boundaries of the two types. By directing at insufficiency of the above schemes, a self-adaptive system of shot boundary detection based on neighbor frame difference and non-neighbor frame difference is realized.

In view of accuracy and detection efficiency of boundary detection, we adopt HSV histogram characteristic value to detect the shot boundary. HSV model is a color model that meets characters of human visual perception, so measurement based on HSV color

space can approximate the feeling of human eyes better. When each frame of image is divided into 3*3 sub-blocks unequally via golden partitioning principle, statistics should be made for HSV histogram of every sub-block and histogram difference of corresponding interval needs to be calculated. Moreover, different weighting coefficients must be set for different intervals. The average distance after weighting will be treated as difference value between two frames of images. Inter-frame difference value calculated in this way can better reflect the difference among frame images of different shots. A self-adaptive local threshold value method is adopted in this algorithm, which has solved selection problems of global threshold value. By computing neighbor frame difference, this method can effectively detect abrupt changes. By computing non-neighbor frame difference, gradual changes can be effectively detected. Integrating the two methods, it can almost detect all shot boundaries no matter what the shot change type is. This method does not need to set global threshold and it can simultaneously detect abrupt changes and gradual changes by the comprehensive judgment strategies, the innovation exist here. This method has high adaptability with low computation cost. Experimental results show it is adaptive and efficient.

2. System Implementation

2.1 System Structure

See the following figure for the structure of self-adaptive shot boundary detection system based on neighbor frame difference and non-neighbor frame difference.

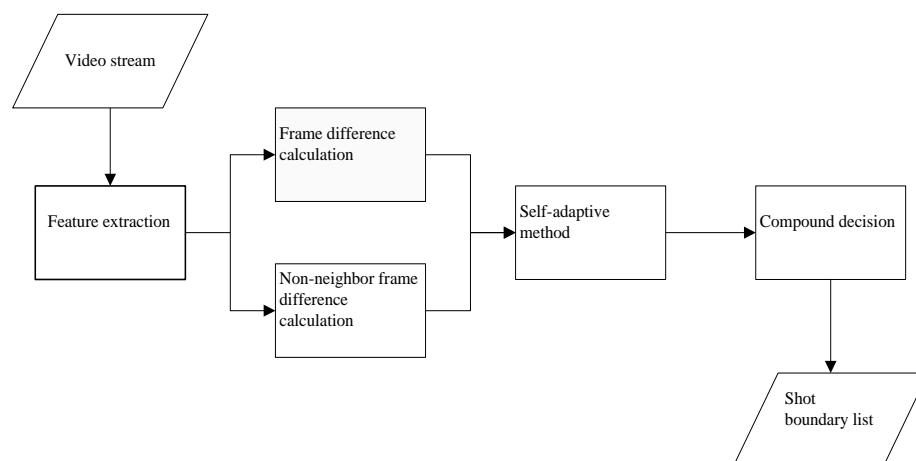


Figure 1. Structure of Shot Boundary Detection System

The existence of shot change boundary has reflected discontinuity of video stream in time, so the selected characteristics should better express such discontinuity. Characteristics applied to shot boundary detection at present mainly include: histogram characteristic value, edge (outline) characteristic value, motion characteristic value, *etc.* Edge characteristic value often calculates edge of the object in the image at first, and then judges shot change by analyzing variable quantity of the edge. Frequently-used edge calculation methods cover Roberts operator, Gaussian Laplacian operator, Prewitt operator, Robinson operator, Kirsch operator, Sobel operator, and Canny operator. The concept of edge change rate proposed in literature [4,7] utilizes such characteristic for abrupt change detection of shot. However, such algorithm requires huge calculated amount, so it is only used for detection of the non-real-time gradual change shot, rather than detection of abrupt change shot. Calculation methods of motion characteristic value include brightness absolute frame difference method based on pixel, method based on

light stream, and residual method based on motion vector or after block matching. However, detection effect of these methods based on motion characteristic value is not superior to histogram method.

Based on features of the above characteristic value selection methods, in view of accuracy and detection efficiency of boundary detection, we adopt HSV histogram characteristic value to detect the shot boundary. HSV model is a color model that meets characters of human visual perception, so measurement based on HSV color space can approximate the feeling of human eyes better. After HSV space model is selected, unequal interval quantification is conducted for HSV color space, so as to cut down dimensions of histogram vectors and reduce the calculated amount [8]. By analyzing the color model, the space of hue H is divided into 8 parts, and the spaces of saturation S and brightness V are divided into 3 parts respectively. Meanwhile, 72 common color grades are obtained after quantification according to different color scopes. Hue, saturation and brightness values after quantification are H, S and V separately [6].

Histogram characteristic value cannot record location information about pixel point, so it's hard to reflect space information about the video image. Literature [3,6] put forward the concept of partitioning histogram, and the entire video frame is equally divided into $N \times N$ sub-blocks. Histogram characteristic value of each sub-block is calculated separately, and then the frame difference is worked out. Symmetrical partitioning for video frame has considered location information about pixel, but it neglects the importance of each block. The scheme of [9] takes the importance of each sub-block into consideration, but such method is similar to Gaussian weighting method, as shown in Figure. 2(a). Experiment shows that such weighting method does not have a good effect for shot detection. In many video programs, advertisement contents, titles or captions often appear above or under the picture; besides, camera shake will change contents around the shot, but no shot change happens at this time. As a result, symmetrical partitioning or simple Gaussian weighting method might cause false shot boundary detection. In order to solve the defects of symmetrical partitioning, length and width of the image will be divided according to the proportion of 3:5:3, rather than 1:1:1, as shown in Figure. 2(b). Such partitioning scheme has reflected change of main contents in the video well. To eliminate the interference of advertisement contents or captions at the top and bottom of the screen, the simple weighting method is modified. For comparison, the modified weighting method and simple weighting method are put together, as shown by Figures in Figure. 2(b). Experiment proves that this partitioning and weighting scheme gains a good detection effect.

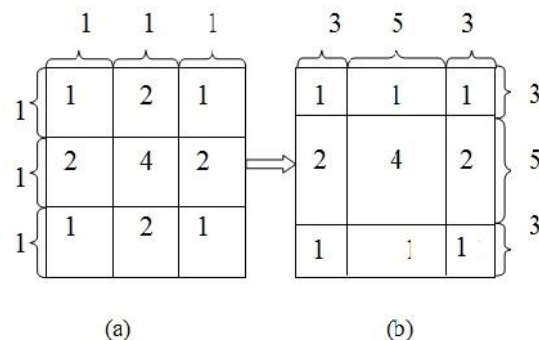


Figure 2. Asymmetrical Partitioning and Weighting Principle

2.3. Calculation of Neighbor Frame Difference and Non-Neighbor Frame Difference

When each frame of image is divided into 3*3 sub-blocks unequally via golden partitioning principle, statistics should be made for HSV histogram of every sub-block and histogram difference of corresponding interval needs to be calculated. Moreover,

different weighting coefficients must be set for different intervals. The average distance after weighting will be treated as difference value between two frames of images. Inter-frame difference value calculated in this way can better reflect the difference among frame images of different shots.

1) Neighbor frame difference can be expressed by weighted sum of histogram distance among corresponding sub-blocks. The distance of each sub-block is

$$d(H_i, H_j) = \sum_{l=0}^n |h_i(l) - h_j(l)|, \quad (1)$$

In the formula, \mathbf{l} represents quantification series of color, and it is equal to 72 in this paper. Suppose that the weighting matrix is

$$W = \begin{bmatrix} w_1 & w_2 & w_3 \\ w_4 & w_5 & w_6 \\ w_7 & w_8 & w_9 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 4 & 2 \\ 1 & 1 & 1 \end{bmatrix}. \quad (2)$$

In the formula, position of weighted value is corresponding to position of partitioning interval. The size and position of partitioning interval can reflect the importance degree of each partitioning interval. Histogram difference value of the interval corresponding to two neighbor frames can be calculated, recorded as d_1, d_2, \dots, d_9 . Then the neighbor frame difference will be gained according to weighting coefficient and interval histogram difference value.

$$D = \sum_{i=1}^9 w_i d_i / \sum_{i=1}^9 w_i, \quad (3)$$

2) Calculation of non-neighbor frame difference

The whole transformation process of gradual change shot is continuous and no great changes will happen to characteristics of neighbor frames. Therefore, we cannot identify the shot through neighbor frame difference. In order to better detect gradual change, a video boundary detection method based on non-neighbor frame difference is proposed. The so-called non-neighbor frame difference means to directly calculate the frame difference between the current frame and number $N(N > 1)$ frame after the current one.

$$D_{ij} = |F_i - F_j| \text{ where } |j - i| = N, \quad (4)$$

In the formula, D_{ij} means the non-neighbor frame difference between number i frame and number j frame; $|F_i - F_j|$ denotes distance between the corresponding frames. Within a shot, content change between both neighbor frames and non-neighbor frames will not be so large. However, when a gradual change happens, contents between neighbor frames is small, but the current shot will gradually turn into another shot and great change will occur to contents of the two shots. Thereby, the non-neighbor frame difference will be great. As a result, by calculating non-neighbor frame difference, the specific gradual change effect can be eliminated and the actual change of shot contents will be reflected. In this way, gradual change shot boundary can be better detected.

The specific calculation method of non-neighbor frame difference is the same with the calculation method of neighbor frame difference. See formula (1), formula (2) and formula (3) in the above.

2.4. Shot Boundary Detection

Self-adaptive method: Under normal conditions, there will be no great difference among different image frames. The inter-frame difference value is in a relatively stable scope, which means that it will fluctuate around a certain mean value. When shot change happens, such stability will be broken. Therefore, the algorithm only utilizes inter-frame difference value information within a limited length before the current frame when identifying shot change. The length is often from the 1st frame after the former shot ends to the 1st frame of the current frame. It can be thought as a sliding window with changeable length. Mean value of inter-frame difference values in this window will be calculated. If the difference between the current inter-frame difference value and this mean value reaches a certain coefficient, the current frame will be thought as the initial frame of the next shot. Local threshold value method is adopted in this algorithm, which has solved selection problems of global threshold value.

(1) Self-adaptive abrupt change shot detection based on neighbor frame difference

Abrupt change of shot mainly happens between two frames. Self-adaptive method is adopted to detect abrupt change shot by calculating neighbor frame difference on the basis of the above method. Sign convention in the algorithm is as follows:

α is the predetermined shot boundary detection threshold coefficient, and the statistical result shows that α should be $5 \cdot 6^{[1]}$;

D_i represents the frame difference between two neighbor frames of video stream;

Mean denotes the mean value of inter-frame difference values in the window;

Sum indicates the cumulative sum of neighbor frame differences;

Count signifies cumulative frequency of frame differences.

Under some special situations, the phenomenon of static shot or almost static shot will appear. At this time, inter-frame difference value is quite small. If the minimum difference value is taken into account in calculation for mean value of inter-frame difference in the window, the mean value will be obviously decreased. An extreme situation is that when there are too many static (or almost static) frames in the shot, mean value of inter-frame difference in the window will tend to be 0. At this time, when the above method is used, erroneous judgment might be caused. In order to avoid this phenomenon, the above judgment method is improved. A minimum threshold value Value is established, and it is taken as 10^{-4} in the algorithm. Before the judgment, D_i is compared with Value. If $D_i \leq \text{Value}$, shot change identification will not be conducted.

In conclusion, description about the algorithm flow is as follows:

1. Initiate the algorithm. Read in the two neighbor frames of video stream, starting from the first and second frames. If the last frame is read, the process should be stopped.

2. Transform two neighbor frames into HSV model according to the method illustrated in [3], and quantify them into 72 color spaces.

3. Calculate the difference value BlkDiff of partitioning histogram and partitioning histograms corresponding to the two neighbor frames; work out the difference value D_i between histograms of two neighbor frames.

4. Judge whether D_i is smaller than the minimum threshold value Value; if it is smaller, shot change identification will not be conducted.

5. Give a definition, to make flag=0 mean the start of a new shot. If a new shot is started, $\text{Sum} = D_i$, $\text{Count} = \text{Count} + 1$, $\text{Mean} = D_i$, and flag=1. Go to ①.

6. Judge whether D_i is greater than α times of Mean. If it is greater, it will be judged as a new shot boundary; mean=0, count=0, and flag=0. Go to ①.

7. When D_i is smaller than α times of Mean, $Sum = Sum + D_i$, Count ++, and $Mean = Sum / Count$. Go to ①.

(2) Self-adaptive gradual change shot detection based on non-neighbor frame difference

In order to realize gradual change detection, self-adaptive detection algorithm based on non-neighbor frame difference is adopted. Sign convention in the algorithm is as follows:

α is the predetermined shot boundary detection threshold coefficient;

D_{ij} represents the frame difference between two non-neighbor frames of video stream (according to the experiment, inter-frame distance of non-neighbor frames should be 6 in this algorithm);

M_n denotes the local cumulative mean value of non-neighbor frame differences;

M_x signifies the maximum value of non-neighbor frame differences in the window;

Mean indicates the mean value of inter-frame distance in the window;

Sum represents the cumulative sum of neighbor frame differences;

Count signifies cumulative frequency of frame differences.

Description about the algorithm flow is as follows:

1. Initiate the algorithm. Set up a sliding window, and 3 non-neighbor frame differences are treated as the window scope. In this way, width of one window is 18 frames, and generally gradual change shot boundary won't be missed. The threshold coefficient of gradual change shot boundary detection is set as $\alpha=3$.

2. Calculate the mean value M_n of non-neighbor frame differences in the first window of new shot.

3. Slide a window backward and calculate M_x in the current window. If $M_x < \alpha M_n$, re-calculate of all windows including the current window and those after the new shot begins. Slide a window backward.

4. If $M_x > \alpha M_n$, and M_x is located in the first half part of the window, the location of M_x is the gradual change position, and the next window should be set as a new shot boundary. Go to 2.

5. Otherwise, position of M_x in the current window should be recorded and M_x of the next window should be calculated. If this M_x is greater than M_x of the former window, this M_x is the gradual change position. Mark this position and set the next window as the new shot boundary. Go to 2. Otherwise, position of M_x in the former window should be the gradual change position. Mark this position. The last half part of the current window is set as content of the next shot. Calculate the mean value of non-neighbor frame differences. Go to 3.

Purpose of step 4 is to consider the situation in which the first half part of the window glides across the gradual change shot boundary; step 5 thinks about the situation in which the last half part of the former window glides across the initial part of a relatively long gradual change shot boundary. Experiment shows that this algorithm gains a good detection effect for gradual change shot boundary.

2.5. Comprehensive Judgment Scheme

Self-adaptive method based on neighbor frame difference can detect abrupt change shot well, while self-adaptive method based on non-neighbor frame difference can detect gradual change shot well. Dynamic integration of these two can detect all shots well. Therefore, the following decision rules are designed:

$$\begin{aligned}
 S_c &= \{C_1, C_2, \dots, C_N\}, \\
 S_g &= \{G_1, G_2, \dots, G_M\}, \\
 S &= \begin{cases} C_i & C_i \in G_j (j=1, 2, \dots, M) \\ S_c \cup S_g & \text{others} \end{cases} \quad (5)
 \end{aligned}$$

In the formula, S_c is the set of abrupt change shots; S_g represents the set of gradual change shots; S indicates the total shot set; N and M denote the total number of detected abrupt change and gradual change shots.

3. Experimental Result and Analysis

The experimental data mainly come from the following websites: <http://www.open-video.org/> and <http://www.alltheweb.com/>; the video format is MPEG-1, the image size is 352*288, and there are 25 frames every second. The experimental environment is Windows XP operating system and PC machine with internal storage of 3.36G and 2.89GHzCPU. The video segmentation position is compared with accurate position of artificial markers, so as to verify accuracy of the test result. As for evaluation criterion of gradual change shot, there are numerous continuous frames at the transformation part of gradual change shot, so an error within 5 frames is allowable.

Precision ratio and recall ratio are used as two indexes to evaluate detection results of shot change boundary.

$$\text{Precision ratio} = \frac{\text{correct detection quantity}}{\text{correct detection quantity} + \text{false detection quantity}}, \quad (6)$$

$$\text{Recall ratio} = \frac{\text{correct detection quantity}}{\text{correct detection quantity} + \text{leak detection quantity}}, \quad (7)$$

To verify the effectiveness of the proposed method, using this method and the global method, adaptive method, the sliding window method, the second frame difference method, reverse split method to detect different video clips. The film sequence has 10255 frames and the number of shots is 198, they are all abrupt change shots, Table 1 shows the various video shot detection results. The total number of frames of the Scenery sequence is 15253, and the number of shots is 252, they are all gradual change shots, Table 2 shows the various results. The teaching sequence have 31253 frames and 165 shots, including 130 abrupt change shots and 35 gradual change shots, Table 3 shows the results of a variety of methods to detect this sequence. The total number of frames of the football sequences is 23388, and the number of shots is 143, including 119 abrupt change shots and 24 gradual change shots, Table 4 shows the various test results.

Table 1. Shot Detection Results of Different Methods to Film Sequence

Different shot detection methods	Shot	Correct detection quantity	False detection quantity	Leak detection quantity	Recall ratio(%)	Precision ratio(%)
This method	198	193	3	5	97.47	98.46

Global method	198	191	3	7	96.46	98.45
Adaptive method	198	188	5	10	94.95	97.41
Sliding window method	198	192	4	6	96.97	97.96
Second frame difference method	198	189	7	9	95.45	96.43
Reverse split method	198	188	7	10	94.95	96.43

Table 2. Shot Detection Results of Different Methods to Scenery Sequence

Different shot detection methods	Shot	Correct detection quantity	false detection quantity	leak detection quantity	recall ratio(%)	precision ratio(%)
This method	252	231	15	21	91.67	93.90
Global method	252	209	30	43	82.94	87.45
Adaptive method	252	225	21	27	89.29	91.46
Sliding window method	252	221	30	34	87.70	86.67
Second frame difference method	252	221	19	31	87.70	92.08
Reverse split method	252	219	22	33	86.90	90.87

Table 3. Shot Detection Results of Different Methods to Teaching Sequence

Different shot detection methods	Shot	Correct detection quantity	false detection quantity	leak detection quantity	recall ratio(%)	precision ratio(%)
This method	165	160	23	5	96.96	87.43
Global method	165	146	18	19	88.48	89.02
Adaptive method	165	159	15	6	95.00	88.24
Sliding window method	165	158	14	7	95.76	91.86
Second frame difference method	165	155	25	10	93.94	86.11
Reverse split method	165	151	28	14	91.52	84.35

Table 4. Shot Detection Results of Different Methods to Football Sequence

Different shot detection methods	Shot	Correct detection quantity	false detection quantity	leak detection quantity	recall ratio(%)	precision ratio(%)
this method	143	141	21	2	98.60	86.95
global method	143	130	15	13	90.91	89.66
adaptive method	143	137	17	6	95.80	88.96
sliding window method	143	136	18	7	95.10	88.31
second frame difference method	143	133	20	10	93.01	86.93
reverse split method	143	130	26	13	90.91	83.33

Using method of this paper, the system has reached a high precision ratio and recall ratio in terms of film sequence with abrupt change shot and scenery sequence with gradual change shot. As for teaching sequence and football sequence containing both abrupt change shot and gradual change shot, the system has reached extremely high recall ratio. However, two strategies of neighbor frame difference and non-neighbor frame difference are used at the same time, so there are many cumulative shots of false detection, and the precision ratio is low. Shots of false detection mainly come from serious camera shake and fierce change of illumination in the camera. Leak detection is caused by the situation where color difference between the two neighbor shots is too small.

4. Conclusion

According to the characteristics of abrupt change and gradual change, this system has gained a good detection effect for abrupt change shot via self-adaptive method by calculating neighbor frame difference. Meanwhile, a good detection effect has been obtained for gradual change shot via self-adaptive method by calculating non-neighbor frame difference. Dynamic integration of these two has acquired a good comprehensive detection effect for both abrupt change shot and gradual change shot. However, due to complexity of video, the system cannot gain an optimum detection effect for all videos. In order to enhance the detection effect of the system, other characteristics including motion characteristic, edge change characteristic and brightness variation characteristic can be added, to improve detection performance of the system. In addition, compared with RGB histogram characteristic, HSV partitioning histogram system has a better detection performance, but its detection efficiency is low. This is a problem that needs to be solved at the next step of our work.

Acknowledgments

This work is supported by project of Shandong Province Science and Technology Development Program (2008GG30001007), China, and by Science and Technology Project of the City of Jinan (201202184), China.

References

- [1] Nagasaka A. and Tanaka Y., "Automatic video indexing and full-motion search for object appearance," Proceedings of the Second Working Conference on Visual Database System, North Holland: Kluwer Academic Publishers, (1991), pp. 113-127.
- [2] Qianlei L., lvxi Y. and Cairong Z., "Second difference and the second difference pixel matching method for video shot change detection," Journal of image and graphics, vol. 2 no. 8, (2003), pp. 161-168.
- [3] Yihua Z., Yuanda C. and Hong X. Z., "A method of abrupt shot detection based on the second frame difference," Computer engineering and applications, no. 6, (2005), pp. 22-25.
- [4] Y. Geng, J. Chen and K. Pahlavan, "Motion detection using RF signals for the first responder in emergency operations," A PHASER project, 2013 IEEE 24th International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), London, Britain September. (2013).
- [5] Huang Z., Wang L. and Shen H. T., "Online near-duplicate video clip detection and retrieval: an accurate and fast system," Proceedings of IEEE International Conference on Data Engineering (ICDE'2009), (2009), pp. 1511-1514.
- [6] Yueting Z., Yunhe P. and Fei W., "Online multimedia information analysis and retrieval," Tsinghua university press, (2002).
- [7] Zhang H. J. and Smoliar S. W., "Developing Power Tools for Video Indexing and Retrieval," SPIE Conference on Storage and Retrieval for Image and Video Databases, San Jose, CA, (1994), pp. 140-149.
- [8] Peng B. and Li B., "A causal abrupt shot detection method," Computer engineering and applications, vol. 5 no. 40, (2004), pp. 91-93.
- [9] Mauro C., Rodrigo D. O. and Nuria O., "Understanding near duplicate videos: a user-centric approach," Proceedings of Conference on Multimedia: the 17th Annual ACM International (MM'2009), Beijing, China, (2009).
- [10] Li X. W., Li Z. M. and Zhang M. X., "Technology of content-based shot detection," Video Engineering, vol. 32 no. 3, (2008), pp. 19-21.
- [11] Wu T. S., Lin H. Y. and Hu W. C., "Audio watermarking scheme with dynamic adjustment in mute period," Expert Systems with Applications, vol. 38 no. 6, (2011), pp. 6787-6792.
- [12] Wang X. Y., Niu P. and Lu M. Y., "A robust digital audio watermarking scheme using wavelet moment invariance," The Journal of Systems and Software, vol. 84 no. 8, (2011), pp. 1408-1421.
- [13] Lei B. Y., Soon I. Y. and Li Z., "Blind and robust audio watermarking scheme based on SVD-DCT," Signal Processing, vol. 91 no. 8, (2011), pp. 1973-1984.
- [14] Sanjay R. A. and Balasubramanian R. A., "A chaotic system based fragile watermarking scheme for image tamper detection," International Journal of Electronics and Communications, no. 1, (2011), pp. 840-847.
- [15] Fan M. and Wang H., "Chaos-based discrete fractional sine transform domain audio watermarking scheme," Computers and Electrical Engineering, vol. 35 no. 3, (2009), pp. 506-516.
- [16] Zhong Y., Karu K. and Jain A. K., "Locating text in complex color images," IEEE Document Analysis and Recognition, no. 1, (1995), pp. 146-149.
- [17] Y. Geng, J. He, K. Pahlavan, "Modeling the Effect of Human Body on TOA Based Indoor Human Tracking," International Journal of Wireless Information Networks, vol. 20 no. 4, pp. 306-317
- [18] Xiang S. and Huang J., "Histogram-based audio watermarking against timescale modification and cropping attacks," IEEE Trans Multimedia, vol. 9 no. 7, (2007), pp. 1357-1372.

Author



Junxiao Liu, was born in Shandong, China. She received the bachelor and master degree in educational technology in July 2002 and July 2007 from Shandong Normal University. Her current research interests are the content analysis of video. Besides, she has begun to do some research on the content analysis of flash.

She worked at Shandong Normal University after her graduation. She has participated in some projects about multimedia researches