International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol. 10, No. 7 (2017), pp.85-92
http//dx.doi.org/10.14257/ijsip.2017.10.7.08

Extraction Algorithm on Representative Frame Images of the Flash Animation Visual Scene

Shi Lin[1, 2] and Meng Xiangzeng[1*]

[1]School of Communication, Shandong Normal University, Jinan, China
[2]Business School, Shandong Jianzhu University, Jinan, China
[1]E-mail: shilin_2010@126.com, [2]mxz_sdnu@126.com

Abstract

As a kind of important network resources, Flash animation has been widely studied, but its network retrieval has a lot of problems. In order to improve the retrieval accuracy and fast preview the search results, we mainly study the process of extraction of Flash animation visual scene representative frame. We proposed an adaptive threshold segmentation algorithm of the Flash animation visual scene, and extraction algorithm of representative frame based on the average main color distance of the key frame. We download 82 Flash animation samples for visual scene segmentation from the Internet, and then we use the correct visual scenes to extract their representative frame. Finally, we follow the classification of MTV, games, cartoons, and advertising to analyze the segmentation results of each type of Flash animation visual scenes and the representative frame extraction effect. The visual scene representation frame obtained in this study can be used to make the dynamic summary of Flash animation in order to preview the search results of network Flash resource search, and our research results have important implications for the Flash resource retrieval based on content.

Keywords: *Visual scene; Representative frame; K-means clustering; Flash retrieval*

1. Introduction

Flash animation has a full application in many fields, and there are a huge amount of Flash resources on the network. However, this is a great challenge to the Flash demanders. So, how to find the required Flash resources quickly and accurately Puts forward higher requirements to the Flash search engine. At present, the Flash retrieval algorithm based on keywords, context information and metadata is relatively mature. But the retrieval efficiency and accuracy are generally not high. So we started the analysis and research on the visual features of Flash animation. We segment the animation to several visual scenes and extract their visual content features, finish the automatic annotation and indexing of the Web Flash resources by analyzing the file structure of Flash animation, and use it to improve the efficiency and accuracy of the Flash resource retrieval.

The visual scene of Flash animation is defined as animation clips which are composed of a series of consecutive frames with similar visual features.

The representative frame is a static image, which can reflect the representative meaning of the visual scene. It is generally able to represent the production intents of the animation producers. The extracting target of a representative frame is to represent the content of a visual scene with a still image. We can build the index of the visual scene to achieve the retrieval of Flash animation by extracting the content feature of the representative frame image. We can also use representation frames of the visual scenes to create animation summary of the Flash. Using the summary, we can make a quick view of the retrieval results of Flash animation resources on the web.

* Meng Xiangzeng

ISSN: 2005-4254 IJSIP

The extracting of visual scene and representation frame image is one of the important contents of Flash retrieval based on content. It will directly affect the retrieval results of Flash animation. In this paper, we mainly study how to segment the Flash to visual scenes effectively, and how to extract the representative frame of the visual scene by using the main color distance algorithm.

2. Related Research

Although Flash animation is the most popular multimedia format in many areas, but the analysis of it is very little because of its complex file structure and the various dynamic effects. Foreign representative researchers such as Dr. Yang Jun of the Carnegie Mellon University proposed a FLAME framework based on the content of the Flash animation retrieval. However, FLAME only defines the framework of the construction of Flash retrieval system, and it can not be directly used for Flash resource retrieval. Dr. Yang has no further study of the content analysis of Flash animation.

Most of the researches of Flash animation content in China are in view of video and image processing technology. The visual scene detection is similar to the shot detection in the video processing, and the representation frame extraction is similar to the key frame extraction of the shot. A lot of video and image processing algorithms can be applied to the content analysis of Flash animation.

In China, for the detection of visual scenes, Liu Lei proposed a method based on color histogram difference. The algorithm needs to determine a global threshold through neural network learning, which is used to measure the similarity between two adjacent frames. The algorithm first calculates the color histogram difference between every two adjacent key frames, and then determines if the difference is less than the threshold value. If so, the two key frames belong to the same visual scene. Otherwise, it means there is a big color change between the two frames, and these two frames do not belong to the same visual scene. In order to reflect the spatial differences of the media object color, the algorithm divides the image into uneven regions, and then carries out different weighted operations on the color histogram of each region to calculate the color similarity between two adjacent images. Then achieve the purpose of detecting the visual scene.

At present, the researches on the representative image extraction of the Flash visual scene are relatively few. In literature [2], it determine the position and number of the representative frame of this scene in term of the number and duration of the key frames in one visual scene, and usually the middle frame image of the visual scene is taken as the representative frame. The algorithm considers the key frame in the middle of the scene is the main picture. This method is simple and easy to operate, but it is one-sided and the extraction accuracy rate of representative frame is not high. Because in many cases, the producer will also use the start frame or the end frame as the main picture. In literature [1], it extract all the key frames of the Flash animation based on the file structure, and then integrate these key frames together to form an animated summary of the Flash. The algorithm does not extract the representation frame images of the scene, and often forms a large capacity animation summary, which is not easy to network transmission.

Aiming at the above shortage of the study on Flash animation visual scene segmentation and representation frame extraction, this paper proposes an improved adaptive threshold algorithm for visual scene segmentation and a representative frame extraction algorithm based on the average main color distance. Our algorithm is very good to make up for the limitation of the global threshold of the visual scene segmentation in the literature [1], and make up for the one-sided of proposed representative frame extraction algorithm.

3. Adaptive Threshold Detection Algorithm for Visual Scene

The color histogram distance of adjacent frames is generally used to detect the visual scene. According to the production principle of Flash file, Flash animation frame is divided into key frame and general frame. In a Flash animation, the producer usually uses a lot of application of gradient, so some frames are transition frames, which are general frames, automatically generated based on the key frames. The difference between these transition frames is small, and the number of the general frames is much larger than that of key frames. Therefore, most of the comparison of all of the adjacent frames in Flash animation is redundant in the traditional algorithm. Therefore, the time complexity of the algorithm is relatively higher. Based on this analysis, we can first extract the key frames from the Flash animation, and then carry on the clustering detection of the visual scene based on the key frames. Therefore, we can greatly reduce the invalid comparison, and improve the analysis efficiency. Whether the extraction of key frame is accurate or not will directly affect the detection results of the visual scene.

3.1. Key Frame Extraction

The key frame of Flash animation generally refers to the frame which defines the change in the object’s property or the assignment of the action. The key frame and the general frame use the same storage format and definition method, and all use a “ShowFrame” tag to play. Therefore, we need to analyze the “ShowFrame” tags based on the structure of SWF file, and then extract the key frames of the Flash animation.

The key frame controls the dynamic effect and user interaction of the objects in Flash animation. The objects include text, graphics, images, video, audio, *etc.* The control includes five cases as allocation of actions, adding objects, removing objects, shape change, and attribute change ［3］. Among them, the allocation action refers to the addition of the Action script, usually using the “DoAction” or “DoInitAction” tags. It uses the “PlaceObject” tags to add objects to the stage, uses the “RemoveObject” and “RemoveObject2” tags to remove objects from the stage, uses “DefineMorph” tags to define the shape change, and uses “PlaceObject2” and “PlaceObject3” tags to change the object property, including the change of the location, size, color, shape, fill, and so on.

Therefore, we can determine whether the current frame is a key frame by judging whether the frame contains tags such as “DoAction”, “DoInitAction”, “PlaceObject”, “RemoveObject”, “RemoveObject2”, “DefineMorph”, “PlaceObject2”, “PlaceObject3”, *etc.* If it is, we should record the frame number of it, and save it as a BMP image file, which will be used for the later visual scene detection and key frame clustering.

3.2. Visual Scene Detection

In literature [1], the algorithm considers the color difference in space to judge the visual scene boundary of the Flash animation, and can achieve a certain effect, but the algorithm depends on the selection of the judge threshold. On the one hand, if the threshold is too large, the frames with different visual features will be classified into the same visual scene, which we call a leak detection. On the other hand, if the threshold is too small, the frames with similar visual features will be classified into the different visual scene, which we call a false detection. Aiming at the shortcoming of the fixed threshold segmentation algorithm, we propose an adaptive threshold segmentation algorithm.

In the same visual scene, the color histogram difference between the key frames should be relatively uniform, and the color histogram difference between any two frames should be similar to the average difference of the whole visual scene. Therefore, we believe that if the difference between a certain frame and its previous frame has a sharp fluctuation, and the difference is much larger than the average difference of the current visual scene, this frame is the scene boundary. It is to say that the frame is the end of the previous visual scene and the start of the next visual scene. Therefore, we can detect the visual scene boundary by judging the size relationship between the color histogram difference of adjacent frames and the average difference of the visual scene.

We obtain a coefficient by neural network learning, called α . The average color histogram of the visual scene is multiplied by α as the threshold value. Then, we think the current frame is the beginning of the next visual scene if the difference between the current frame and the previous frame is larger than the threshold of the visual scene the previous frame belongs. In addition, the previous frame is the end frame of the last visual scene. Otherwise, the current frame and the previous frame belong to the same visual scene. Then we continue to judge the next frame in the same way. It can be seen that the judgment threshold of each visual scene is not the same, and the threshold changes with the variation of the inter frame difference.

We determine the adaptive threshold for each visual scene by the following formula.

$$T_k = \frac{\sum D(h_i, h_{i+1})}{n} \times \alpha \tag{1}$$

In formula (1), T_k is the adaptive threshold for the kth visual scene, and α is the coefficient that acquired by machine learning. $\sum D(h_i, h_{i+1})$ is the sum of the difference between the adjacent frames in the previous visual scene. n is the total number of frames in the previous visual scene.

We still take the weighted algorithm of color histogram difference in literature [1] as the algorithm for our calculation of the frame difference.

Based on the color histogram distance and the adaptive threshold, we detect the edge of the visual scene by the following steps.

Step 1, extract the key frames of the Flash animation using the tag algorithm, put the frame number of the key frame into the array $A[]$, and store the key frames in the BMP image format.

Step 2, analyze the image obtained by the last step, calculate the color histogram difference between adjacent frames, and store the differences into the array $B[]$.

Step 3, initialize the variables, like $n = 1$, $i = 2$, $K[1] = 1$. n represents the number of visual scenes, and $K[n]$ represents the start frame number of the nth visual scene.

Step 4, calculate the sum of the color differences between the frames in the nth visual scene. If $B[i] \geq Sum$, then $n = n + 1$, $K[n] = i + 1$, $i = i + 1$, continue to perform step 4. Otherwise, there is no new scene, then $i = i + 1$.

Step 5, judge the variable i. If i is less than or equal to the total number of key frames, then continue to perform step 4. Otherwise, exit the program.

4. Algorithm of Representative Frame Extraction

Studying the animation production principle and skills, we know that the producers tend to use a specific hue in a scene to express their production intention, such as the use of green, blue and so on to describe the quiet, cold, fresh feeling, and the use of red, orange, yellow and other colors to describe the warm, intense, joyful feeling.. Therefore, we study the extraction algorithm of the representative image in a visual scene. That is to first calculate the main color value of the visual scene, and then choose the key frame whose hue is most similar to the main color value as the representative frame of the scene. The algorithm steps are as follows.

Step 1, get the main color of the visual scene. We take the average value of all key frames in the scene as the main color value. The main color of the key frame is the color value of the largest number of prime numbers in the frame.

Step 2, calculate the Euclidean distance between the main color value of each key frame and the main color value of the scene, and select the key frame with the smallest distance as the representative frame of the visual scene. The Euclidean distance formula is as follows.

$$\mathrm{d} = \sqrt{(R_1 - R_2)^2 + (G_1 - G_2)^2 + (B_1 - B_2)^2} \tag{2}$$

In this paper, the flow chart of the visual scene detection and representative frame extraction is as Figure 1.

5. Experimental Results

In order to verify the validity of this algorithm, we downloaded 82 Flash experimental data from Internet include MTV, games, cartoons and advertising, *etc.* First, we marked the visual scene and representative frames of the Flash animation artificially, and then set up the sample database, as a measure of the accuracy of the algorithm. We got 1791 visual scenes with the algorithm, and the experimental data as shown in Table 1.

We use the segmentation accuracy to evaluate each class of samples, and analyze the segmentation effect of the adaptive threshold algorithm. The segmentation accuracy is equal to the number of correctly visual scenes obtained through the algorithm divided by the total number of visual scenes obtained manually.

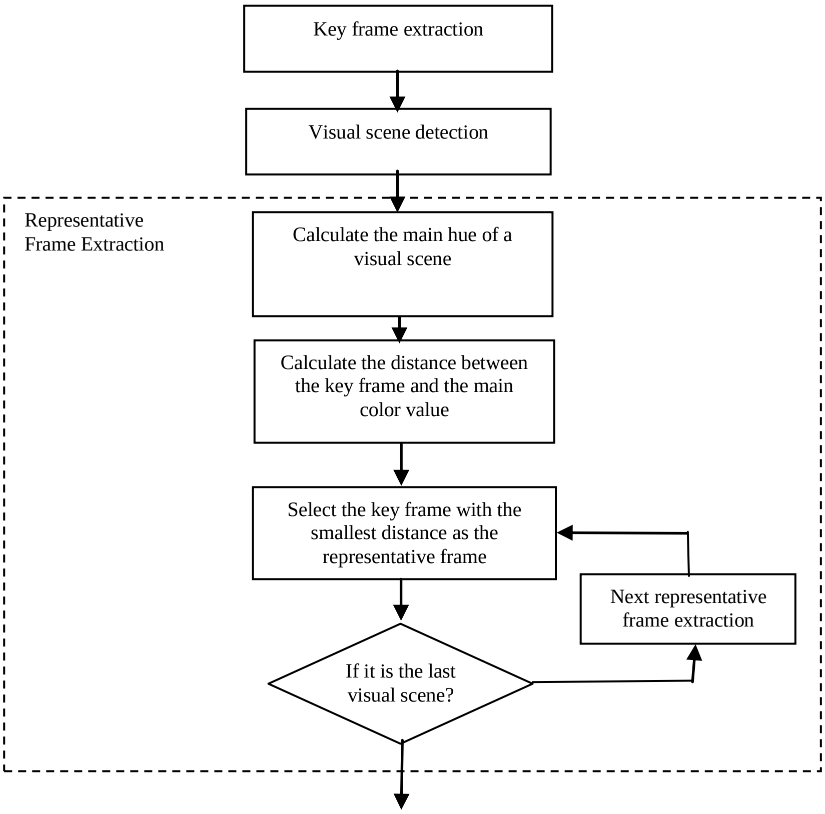
Key frame extraction
Visual scene detection
Representative Frame Extraction
Calculate the main hue of a visual scene
Calculate the distance between the key frame and the main color value
Select the key frame with the smallest distance as the representative frame
Next representative frame extraction
If it is the last visual scene?

Figure 1. The Flow Chart of the Representative Frame Extraction

Table 1. The Visual Scene Detection Data

Flash Type	Number of animation	The number of visual scenes automatically segmented	The number of visual scenes artificially segmented	Accuracy rate of visual scene segmented
MTV	22	764	704	86%
games	20	91	109	57%
cartoon	20	698	674	77%
advertisement	20	238	173	58%

For the detection of the visual scene, the accuracy of the algorithm mentioned in this paper is higher for MTV and cartoon. Because the change of these two kinds of Flash scene is slow, and there is a clear frame switching effect, artificial detection and automatic detection can carry out smoothly. However, because of its rapid change, high frequency and many scenes, it is difficult to define the boundaries of the scenes in games and advertisement. Therefore, it affects the accuracy of its visual scene segmentation. However, the visual scene segmentation algorithm in this paper is able to have a good scene segmentation effect on various types of Flash animation. For the representative frame is correct or not, we judge it using the similarity of the representative frame extracted automatic by our algorithm and the representative frame in the sample library.

Based on the segmentation of Flash visual scene, we extract the representative frame from the 1287 accurately detected visual scenes. We adopt two algorithms, the algorithm in literature [2] and the algorithm in this paper, to extract the representative frame from the visual scene of the data set mentioned above. In literature [2], the algorithm is to extract the middle frame as a representative frame of a visual scene. In order to test the accuracy of the representative frame, we also apply the method of artificial marking to the representative frames. We use the result data of artificial marked as the measure of the accuracy of automatic extraction. The experimental data is shown in Table 2.

Table 2. Experimental Data of Representative Frame Extraction

Flash type	Extraction accuracy of the algorithm in literature [2]	Extraction accuracy of the algorithm in this paper
MTV	45%	78%
games	75%	88%
cartoon	52%	81%
advertisement	40%	68%

Part of the extracted pictures are as follows, in which Figure 2 is the key frame sequence of the visual scene; Figure 3 is the representative frame extracted using algorithm in literature [2]; and Figure 4 is the representative frame extracted using algorithm in this paper. Space is limited, so we only take two extracted representative frames of the visual scene.

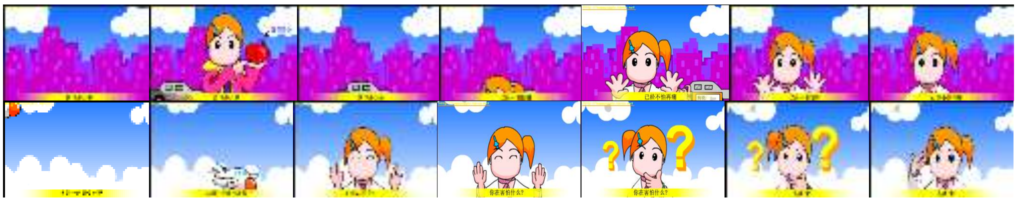

Figure 2. Two Frame Sequence of the Visual Scene

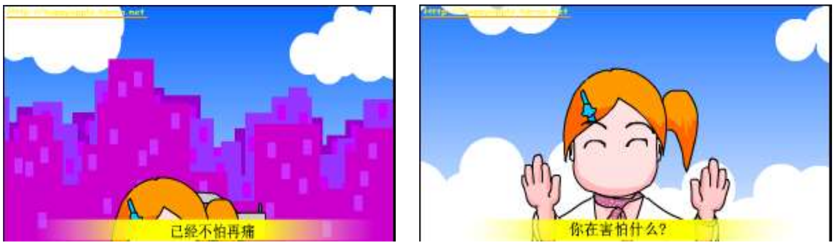

Figure 3. The Representative Frames Extracted Using the Algorithm in Literature [2]

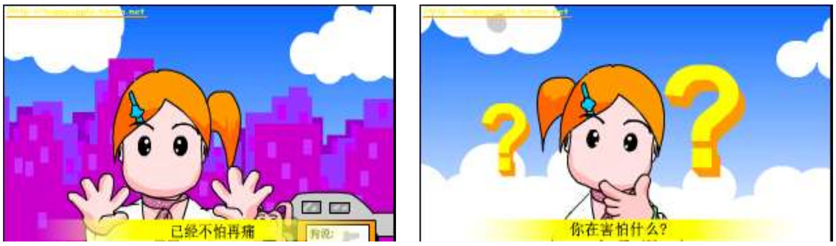

Figure 4. The Representative Frames Extracted Using the Algorithm in this Paper

To analyze the experimental results, the algorithm in this paper gives a better extraction result for the representative frames of various types of Flash visual scenes. The extraction accuracy of MTV and cartoon using the algorithm in literature [2] is not high, because the creation characteristics of these two kinds of animation determine the main frame usually at the beginning, middle and end of the scene. In addition, the extraction accuracy of advertisement is the lowest. Analyzing the reason ,we found that the main frame of the Flash advertisement is usually at the end. As can be seen from the figure, the method of extracting middle frame as representative frame in literature [2] is one-sided for some of the visual scene. On behalf of the game animation, the accuracy rate is higher, because the overall background change of the game animation is small, and the middle frame can basically represent the main content of a visual scene. In this paper, we eliminate the one-sided of the algorithm in literature [2], and give a better extraction result for various types of Flash visual scene. However, if you want to extract more than one representative frame from a visual scene, our algorithm needs to be improved. We cannot simply extract two of the most close to the main color of the scene, so that it is easy to get a repeat of the representative frame. For this problem, we can consider using k-means clustering algorithm to achieve the extraction of multiple representative frames.

6. Summary

Flash animation has become hugely popular around the world because of its compact, easy production, strong interaction and other advantages. In a short time, the network has produced a huge amount of Flash animation resources, which makes the Flash animation retrieval demand higher and higher. The visual scene is a frame sequence of adjacent frames with similar visual features in a Flash animation. The representative frame refers to the key frame, which can mostly reflect the author’s intention, in a visual scene. The visual scene segmentation and the representative frame extraction are important parts of the Flash animation retrieval based on the content. We can extract the visual scenes and the representative frames from a Flash animation. Then compress the representative frames into the dynamic summary of the Flash animation, which is used to a quick browse of the research result, and is more user-friendly, and can greatly improve the retrieval efficiency and accuracy. In this paper, we firstly segment visual scenes from the Flash animation based on the adaptive threshold, and then extract the representative frames of every visual scene based on the main color of the visual scene using the color distance. Experimental results show that the proposed algorithm can extract the representative frame of visual scene very well from Flash animation such as MTV, games, cartoon, advertisement, *etc.* The algorithm is simple and efficient; the accuracy is high; and it can effectively avoid the one-sided of only extract representative frame at the beginning, middle and end of the scene. Therefore, the algorithm has good generality. In the following research, we can consider to study the effect of edge density method for visual scene segmentation, and use the k-means algorithm to extract multiple representative frames.

References

[1] L. Liu, Q. Ding and X. Meng, "Study on feature extraction of Flash movie", China Educational Technology, vol. 9, **(2007)**, pp. 103-106.

[2] F. Liu and X. Meng, "Study on content features extraction and image information analysis of Flash movie", Modern Educational Technology, vol. 12, **(2009)**, pp. 91-94.

[3] D. Ding, J. Yang and Q. Li, "What can Expressive Semantics Tell: Retrieval Model for a Flash-Movie Search Engine", vol.133, **(2005)**, pp. 123-125.

[4] Q. Peng and H. Li, "Video background extraction method based on block histogram analysis", Journal of Southwest Jiaotong University, vol. 1, **(2006)**, pp. 48-56.

[5] X. Meng and L. Lei, "On Retrieval of Flash Animations based on Visual Features", Lecture Notes of Computer Science, vol. 3, **(2007)**, pp. 48-51.

[6] J. R. Smith and S. F. Chang, "Visually Searching the Web for Content", IEEE Multimedia Magazine, vol.4, no. 3. **(1997)**, pp. 12-20.

[7] Y. Rui, T. Huang and S. Chang, "Image retrieval: current techniques, promising directions and open issues", Journal of Visual Communication and Image Representation, vol. 10, **(1999)**, pp. 1-123.

[8] X. Meng, "Analysis of content and structure of network teaching resources", E-education Research, vol. 10, **(2010)**, pp. 81-85.

Authors

Shi Lin, March 10, 1977, Male, Chinese. Working unit: Shandong jianzhu University. Lecturer, postgraduate student, PhD. Research direction: Computer application in Education. Tel: 13153010172. Address: Business School of Shandong Jianzhu University, Jinan City, Shandong province, China