

# Research on Automatic Classification Technology of Flash Animations based on Content Analysis

ZHU Xiao-wei

School of Arts, Harbin University of Science & Technology, Harbin, China

Email: ZhuSmile2012@qq.com

**Abstract**—As a prevailing web media format, Flash animations are delivered and viewed by millions of Internet users every day. The search and classification technologies of Flash animations are important and necessary, but they are difficult due to the absence of understanding of the animation content and are not thoroughly addressed. By mining the file structure and content structure of Flash animations, an automatic analysis method based on the contents of Flash animations is explored, and some typical content features are extracted. These features include file size, graph number, image number, sound number, movie clips number, deformation number, button number, text number, script number, frame number and so on. Using these features, three common classification models, including classification and regression tree, neural network and support vector machine, are respectively selected to effectively classify Flash animations into 5 categories: game, cartoon, MTV, advertisement and teaching courseware. The experimental results show that the neural network model can make full use of various content feature information in Flash animations and has the best classification effect with 90.26% of the average accuracy rate. Moreover, through conducting respectively experiments for the different categories of Flash animations, it is found that game is obviously different from others and they are the most distinguishable category of Flash animations. This research owns important reference value and practical significance in the content analysis, feature extraction, automatic annotation, intelligent search and classification management of Flash animations.

**Index Terms**—Flash Animations; Content Analysis; Feature Extraction; Classification

## I. INTRODUCTION

As a kind of emerging multimedia form, Flash animations are an important element in the Internet. They are widely applied to game, cartoon, MTV, advertisement and teaching courseware, and also appear as the container of texts, PPT and videos in the Internet because they have many advantages of strong artistic expression, simple fabrication, flexible interaction, small documentation, convenience for network communication, etc. The number of Flash animations is on the increase and at present, they have become very important multimedia network information resources [1].

With more and more Flash animations in the Internet, it becomes a noticeable concern that how to manage them enables people to correctly and conveniently find the resources they need [2]. The effective method to solve this problem is to classify them. The efficiency of management and retrieval can be greatly improved through the use of classification technologies.

Presently, there are very limited studies on the classification technology of Flash animations and the studies mainly focus on the classification method based on the external features and contextual information of Flash animations. The external features and contextual information mainly include filenames, keywords, titles and description texts in their webpage, metadata information and creation time of Flash animations. This method is feasible in some degree, but in many cases cannot reveal and express the factual content and semantic relationships of Flash animations for lacking correct understanding of the content of Flash animations [3]. There is one of several representative researches that Jun Yang and Dawei Ding from City University of Hong Kong have proposed a retrieval method based on the content and context of Flash animations including the extraction of various internal object features and events of Flash animations and the analysis of man-machine interaction in Flash animations, and established the classification framework of FLAME and the prototype system based on the content of Flash Movies [4, 5]. The system uses the Bayesian classifying algorithm to classify the Flash Movies, and has achieved some effects, but it lacks systematic theoretical descriptive models and fails to establish and use a whole content management platform [6].

Through the analysis of content features and research on classification algorithms of Flash animations, this paper aims to establish a classification system for Flash animations based on content analysis and achieving the automatic classification management for Flash animations. The second section introduces the file and content formats of Flash animations; the third section mainly describes the extraction methods of various content features in Flash animations; the fourth section discusses the experimental data, experiment and analysis of the experimental results; and the last section gives a summary.

## II. RELATED WORK

### A. File Structure of Flash Animations

The extension of flash animation file is SWF, thus the flash animation file is also called SWF file. SWF file adopts a kind of file structure which is similar to XML file, and it uses a series of binary tags to define different objects as well as their attributes. In the course of analysis, using tags to define files can skip the unrecognized tags for keeping good compatibility of files and helping the extension of files. As a file structure with tags, a SWF file consists of three parts: file header, main part (also called tags) and file closing tag. The storage structure of SWF file is shown in Figure 1.



Figure 1. Storage structure of SWF files

A file header defines some basic information of the SWF file, such as file identification, file version, file size, stage size, frame rate, frame number, etc. The main part of a SWF file defines all of media objects and their features, and uses binary mode to encode its various media objects and their attributes according to certain rules [7]. Therefore, a SWF file cannot be directly understood, but we can work out all of the media objects and their attributes in SWF files through understanding a series of encoding rules of SWF file. The main part of SWF files consists of a series of tags. These tags are used to define different media objects and their attributes, and adopt the same format. Each tag of a SWF file is an independent entity, has no data link with any other tags, and can be removed, inserted and modified. Therefore, the unrecognized tags can be skipped in the process of analyzing SWF files, which has no influence for the analysis of other tags. The closing tag of SWF files is used to label the end of the whole file.

Flash animations are formed by inserting in-between frames between two key frames through a certain interpolation algorithm. This function is implemented by the definition tag and the control tag of the main part of SWF files. Definition tags are used to define objects and their attributes in SWF files, such as the graphs, images, texts, sounds, etc. Definition tags will assign a unique ID to each object as the identification of the object. Control tags will find the objects according to their own ID to control the display, the attribute changes of the objects and the whole play process of the animation. It is SWF files that realize the effect of various animations with the two tags. First, definition tags define a series of objects that are put in an object library, and at the same time assign a unique ID to each object. Control tags search in the object library according to the ID, and when control tags find the object, they will add some actions to the objects such as display, rotation and scaling. Using definition tags or control tags individually cannot draw pictures or form animations, and only each tag coordinating with another one in use can animations be formed. Only when all the three, the definition tag, the

object library and the control tag are used cooperatively with each other can Flash animations be formed.

### B. Content Structure of Flash Animations

In order to describe the content structure of Flash animations better, Flash animations is divided into logic scene and visual scene. The scene in Flash animations which can play automatically and continuously and is made of logically relatively independent frame sequence is called logic scene that can form hyperlinks structure of Flash animations through skip links of key frames. Based on scene clustering and shot detection of videos, when making time lines, adjacent frame sequences with similar visual features are divided into a scene which is called visual scene. Flash animations are made of one or more logic scenes that are laid out in linear array when making time lines and form a nonlinear scene structure through a skip link [8]. Each logic scene also includes one or more visual scenes which can express plenty of visual features. The visual scene includes one or more kinds of the media objects such as graphs, images, sounds, movie clips, texts, etc., and these combine together to form a rich Flash animation. The content structure of Flash animations is shown in Figure 2.

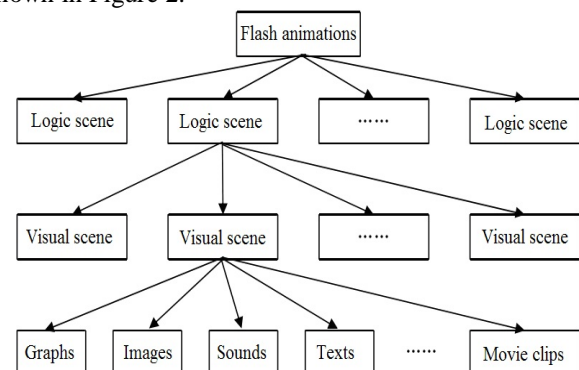


Figure 2. Content structure diagram of flash animations

## III. TECHNICAL METHODS

### A. Extracting the Features of Various Media Objects

Using the method of content analysis, the features of various media objects in Flash animations can be extracted from definition tags by mathematical calculation according to the SWF format description. We can first transform SWF files that are hard for analysis into XML files, and then extract the media objects of graphs, images, sounds, movie clips, etc. and their features in Flash animations through analysis of the XML files. There have been some open source tools, such as Swfmill, Javaswf, etc., that can automatically complete the extraction of content features [9-10]. The general procedures for the feature extraction of various media objects in Flash animations are as follows:

Step 1: Read the content of the first tag in SWF file;

Step 2: Determine the tag types. If it is a definition tag, continue next step; if not, turn to step 5;

Step 3: Adopt different methods of feature extraction according to the media object type defined by the definition tag;

Step 4: Extract corresponding object features as the compositional content information of the Flash animations, and conserve them to the database for later use;

Step 5: Read in the content of next tag, and determine whether it is a closing tag. If it is a closing tag, exit; if not, turn to step 2.

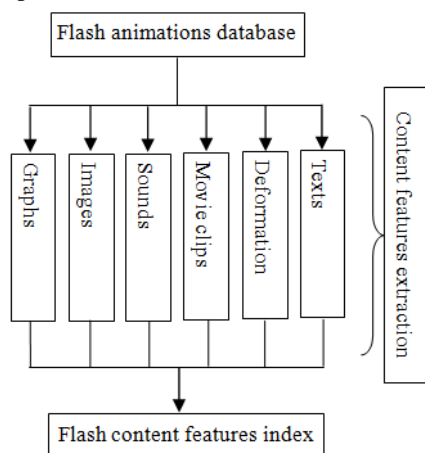


Figure 3. Content Analysis and feature extraction of flash animations

The content analysis and feature extraction of Flash animations include the extraction modules of tag information, visual scene, logic scene and media objects [11]. The extraction module of media objects includes the content analysis and feature extraction of many component elements, such as graphs, images, sounds, movie clips, deformation, buttons, texts, etc., as shown in Figure 3.

The feature extraction of graphs in Flash animations: the number, ID number, minimum enclosing rectangle, fill color of graphs can be extracted through the analysis of the definition tag of DefineShape, etc. in Flash animations; the feature information of the location, movement, rotation and scaling of graphs can be extracted through the analysis of the control tag of PlaceObject. The extractive information of graphs is stored in the database.

The feature extraction of images in Flash animations: the images and the feature and attribute of images, such as, the basic attribute of the size, location, rotation and scaling of images can be extracted through the analysis of the image storage means in Flash animations, and the graphic ID information filled by images is further analyzed [12]. The extractive information of images is stored in the database.

The feature extraction of sounds in Flash animations: the feature and attribute of event audios can be extracted through analyzing the tags of DefineSound, StartSound, and StartSound2 in the definition broadcasting of event audios in Flash animations; the feature and attribute of streaming audios can be analyzed through the analysis of the tags of SoundStreamHead and SoundStreamHead2; the content feature and attribute extracted from sounds

include the number of sounds, ID information, the type of sounds, the sampling rate, bit resolution, soundtrack of event sounds and the broadcasting sampling rate, bit resolution, soundtrack of stream sounds [13]. The extractive information of sounds is stored in the database.

The feature extraction of movie clips in Flash animations: the number, ID number and frame number of movie clips in Flash animations can be extracted through analyzing the definition tag of DefineSprite in the definition movie clips of Flash animations; the size, location, rotation, scaling, movement and color changes of actual movie clips can be extracted through the analysis of the control tag of PlaceObject. The extractive information of movie clips is stored in the database.

The feature extraction of deformation in Flash animations: the number and ID information of deformation in Flash animations can be extracted through the analysis of the definition tag of DefineShapeMorph and DefineShapeMorph2; the movement, rotation, scaling and process information of actual deformation can be extracted through the analysis of the control tag of PlaceObject2 [14]. The extractive information of deformation is stored in the database.

We can finish the extraction of all kinds of content information in Flash animations in the same way, including the features of buttons, texts, scripts, etc., then have a statistical analysis for the extractive content information, and choose the file size and the number of graphs, images, sounds, movie clips, deformation, buttons, texts, scripts, frames, etc. as the feature items for the description of Flash animations and the categorization features of Flash animations. Finally, we can select different classification technology of Flash animations based on content analysis.

### B. Classification based on Content Analysis

Classification is one of the more general problems of pattern recognition, data mining or machine learning, and has been widely studied in various fields. There are many mature classification technology, for example neural network (multi-layer perceptron), naive bayes, gaussian mixture model, support vector machines, k-nearest neighbors, gaussian, decision tree and RBF classifier. In our research, we will only select three typical classifiers for compare, including the classification and regression tree, neural network, and support vector machine.

**Classification and regression tree (CART).** A decision tree is usually used to build a model that predicts the output variable based on several input variables. A typical example is represented in figure 4. Each interior node corresponds to one of the input variables, and there are edges to children for each of the possible values of the input variable. Each leaf represents a value of the output variable given the values of the input variables represented by the path from the root to the leaf. Classification and regression tree is a non-parametric decision tree learning algorithm that produces either classification or regression trees, depending on whether the dependent variable is respectively numeric or categorical.

Decision trees are created by a collection of rules based on variables in the training dataset. Based on input variables' values rules are selected to get the best split to differentiate observations according to the dependent variable. Once a rule is selected and splits a node into two, the same process is applied to each child node. Splitting will continue until CART detects no further gain can be made, or some pre-set stopping rules are met.

Each branch of the tree will end in an output node. Each observation can fall into one and exactly one output node, and each output node is uniquely defined by a collection of rules.

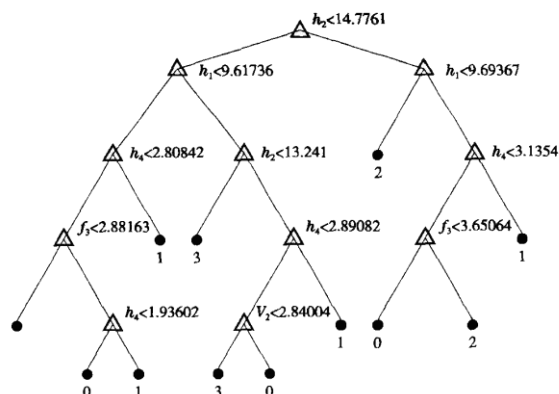


Figure 4. Diagram of decision tree

**Neural network (NN).** An artificial neural network is a mathematical model inspired by biological neural networks and can be used for modeling complex relationships between input variables and output variables or to find patterns in training dataset. A multilayer perceptron (MLP) is a feedforward artificial neural network model that maps sets of input variables onto a set of appropriate output variables. A typical example is represented in figure 5. An MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Each node in one layer connects with a certain weight  $w_{ij,k}$  to every node in the following layer. Where,  $w_{ij,k}$  means the weight of the  $i^{th}$  node in the  $k$  layer to the  $j^{th}$  node in the  $k+1$  layer. Except for the input nodes, each node is a processing element with a nonlinear activation function. MLP utilizes a supervised learning algorithm called back propagation (BP) for training the neural network. MLP is a modification of the standard linear perceptron and can distinguish data that are not linearly separable.

The features extracted from various media objects in flash animations are the inputs of the neural network; the predicted class is the output value of the network.

The number of input nodes corresponds to the number of the features extracted from various media objects in flash animations. Neural network parameters, the weights, are adjusted by the training algorithm according to the least mean square error criterion.

After some preliminary experiments with different network architectures, we choose the two-layer back propagation (BP) neural network with a single linear output node and a hidden layer of log-sigmoid nodes to

estimate the class. We try to vary the number of hidden layer nodes, and the half of input feature dimension is well.

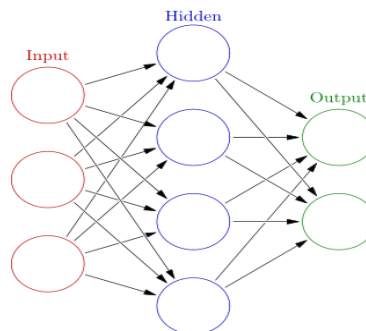


Figure 5. Diagram of neural network

**Support vector machine (SVM).** A support vector machine is supervised learning model that analyze relationships between input data and recognize patterns, and is usually used for classification and regression analysis. The basic SVM takes a set of input variables and predict variables, for each given input variables, which of two possible classes forms the output variable, making it a non-probabilistic binary linear classifier. Given a collection of training samples, each marked as belonging to one of two categories, a SVM training algorithm constructs a model that can assign new samples into one category or the other. A SVM model is a representation of the samples as points in space, mapped so that the samples of the different categories are divided by a clear gap that is as wide as possible. New samples are then mapped into the same space and predicted to belong to a category based on which side of the gap they fall on. A typical example is represented in figure 6.

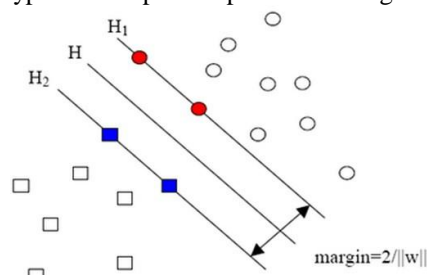


Figure 6. Diagram of support vector machine

TABLE I. PROPORTION OF EACH CATEGORY IN THE DATABASE OF FLASH ANIMATIONS

Categories	Game	Cartoon	MTV	Advertisement	Courseware
Quantity	352	203	102	93	87
Proportion	42.06%	24.25%	12.19%	11.11%	10.39%

A support vector machine creates a hyperplane or set of hyperplanes in a high-dimensional or infinite-dimensional space, which can be used for regression, classification, or other similar tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class. The distance is called functional margin, and the larger the margin is, the lower the generalization error of the classifier is in general.

TABLE II. STATISTICAL ANALYSIS FOR CONTENT FEATURES OF FLASH ANIMATIONS ACCORDING TO THEIR CATEGORIES (ONLY A PART OF CONTENT FEATURES)

Features	Statistical items	Game	Cartoon	MTV	Advertisement	Courseware
File size	Mean value(KB)	1162.21	1783.65	1563.52	314.56	704.58
	Standard Deviation	1864.96	2122.43	1984.23	847.67	963.45
Number of graphs	Mean value(PCS)	386.49	426.19	275.33	51.36	187.44
	Standard Deviation	403.37	364.98	284.25	82.62	190.37
Number of images	Mean value(PCS)	88.26	9.70	9.68	16.82	26.41
	Standard Deviation	92.64	16.97	16.38	34.86	131.43
Number of sounds	Mean value(PCS)	11.37	10.57	1.42	4.65	1.76
	Standard Deviation	20.01	14.89	1.38	4.19	6.94
Number of movie clips	Mean value(PCS)	143.95	47.74	29.82	17.41	26.32
	Standard Deviation	202.49	71.32	38.44	25.76	31.16
Number of deformation	Mean value(PCS)	6.30	12.91	17.88	3.45	4.94
	Standard Deviation	101.23	43.72	32.29	14.12	25.24

When to describe the original problem in a finite dimensional space, it often happens that the data in training dataset are not linearly separable in this space. So, it is a feasible approach to map the original finite-dimensional space into a much higher-dimensional space, presumably making the separation easier in higher-dimensional space. To make the computational load reasonable, the mappings used by SVM schemes are designed to ensure that dot products may be computed easily in terms of the variables in the original space, by defining them in terms of a kernel function  $K(x, y)$ . The hyperplanes in the higher-dimensional space are defined as the set of points whose dot product with a vector in that space is constant. The vectors defining the hyperplanes can be selected to be linear combinations with weights  $\alpha_i$  of feature vectors that occur in the training dataset. In this situation, the points  $x$  in the feature space that are mapped into the hyperplane are defined by the condition:  $\sum_i \alpha_i K(x_i, x) = \text{constant}$ . When

$K(x, y)$  becomes small as  $y$  grows further away from  $x$ , each element in the sum measures the degree of nearness of the test point  $x$  to the corresponding data base point  $x_i$ . So, the sum of kernels above can be used to measure the relative closeness of each test point to the data points originating in one or the other of the sets to be discriminated.

#### IV. EXPERIMENT AND ANALYSIS

##### A. Experimental Database

According to its content and function and with the purpose of its classification and management based on the content analysis, Flash animation can be divided into five common categories: game, cartoon, MTV, advertisement and teaching courseware. We have collected 837 Flash animations from the Internet to establish the database of Flash animations, and manually classify them. The proportion of each category is shown as follows.

It can be seen from Table I that the category of game is in the highest proportion, nearly accounting for the half of samples, which means Flash games in the Internet have a tremendous number and a strong vitality; the category of cartoon is in the second place, accounting for

about the double of each other category; the proportion of MTV, advertisement and teaching courseware is very nearly the same; the proportion of teaching courseware is low.

Through the automatic analysis and extraction of the content of Flash animations, we choose the file size and the number of graphs, images, sounds, movie clips, deformation, buttons, texts, frames, etc. as the description of the features of Flash animations, classify the Flash animations, and have a statistics analysis for these features of all the animations in the database of Flash animations according to their categories, as shown in Table II.

##### B. Experimental Results and Analysis

In this paper, accuracy rate is used to express the effects of Flash classification.  $K$ , the accuracy rate of classification, represents the number of Flash animations classified correctly divided by the total quantity of Flash animations. The formula is as follows:

$$K = \frac{\text{Number of Flash animations classified correctly}}{\text{Total quantity of Flash animations}}$$

We use separately 3 common classifiers, including classification and regression tree, neural network and support vector machine as classification models, and conduct experiments by using the mode of 5-fold cross-validation.

The completion of various classifiers took use of the New Zealand WEKA tools. For the classification and regression tree model, WEKA C4.5 algorithm and its default settings were employed. The neural network model was achieved through the WEKA Multilayer Perception (MLP) using a hidden layer. The number of nodes in the hidden layer was a half of the input feature dimension. The SVM model took use of the WEKA SMO classifier and its default settings.

TABLE III. CLASSIFICATION RESULTS OF FLASH ANIMATIONS

Classification models	Average classification accuracy rate K / %
Classification and regression tree	81.63%
Neural network	90.26%
Support vector machine	87.94%

The final experimental results are shown in Table III. The neural network model has the best classification



effect with 90.26% of the average classification accuracy rate.

The neural network that has the best classification effects is used as classification models, and conduct respectively experiments for the different categories of Flash animations. The classification accuracy rate of each category is obtained, as shown in Table IV.

TABLE IV. CLASSIFICATION RESULTS OF FLASH ANIMATIONS ACCORDING TO THEIR CATEGORIES

Categories	Total quantity	Classification accuracy rate /%
Game	352	92.35%
Cartoon	203	87.93%
MTV	102	85.23%
Advertisement	93	73.52%
Courseware	87	86.15%

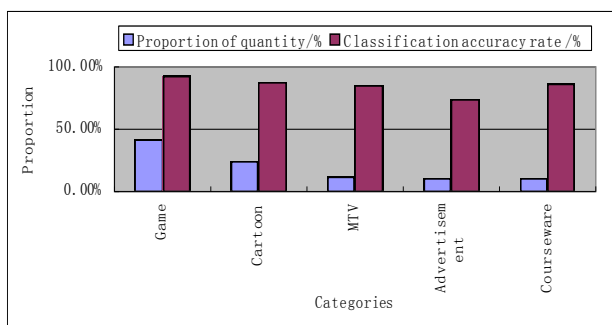


Figure 7. The proportion of quantity and the classification accuracy rate of different classes of flash animations

It can be seen from Figure 7 that the classification method researched in this paper has relatively high overall classification accuracy for Flash animations, especially for the category of game, of which the classification accuracy rate is the highest, and this is because Flash games are extraordinarily rich in content and usually have a large file size, which makes them obviously different from other categories, so it's easy to distinguish them; However, the category of advertisement has the lowest classification accuracy rate, and this is because the category of advertisement is not obvious in interaction and in other features.

## V. CONCLUSION

The content analysis and feature extraction of Flash animations are a research topic that involves very extensive aspects, and have just begun in recent years. In this paper, only the extraction method of the basic content features of Flash animations' internal media resources is discussed, and there are many other aspects needing more and further research, such as the scenario analysis, high level semantic information contained in sounds and images, and interaction analysis of Flash animations, which will become an emphasis in our further research. With the continuous development of multimedia content analysis techniques, more and more researchers will devote themselves into this field, the content analysis and feature extraction of Flash animations will gradually

come to maturity, the classification of Flash animations will be more accurate, and the smart search and content management platform of Flash animations will become more perfect.

## REFERENCES

- [1] Yapei Zhu, Mei Yu, Gunjun Zhang, Zongju Peng, Feng Shao, and Gangyi Jiang, "New Network Bandwidth-limited Multi-view Video plus Depth Coding Method for 3D Video", *Journal of Multimedia*, Vol 8, No 1 (2013), pp. 8-15, Feb 2013 doi: 10.4304/jmm. 8. 1. 8-15.
- [2] WenZhu Sun, HongYu Wang, and DaXing Qian, "A Novel Error Resilient Scheme for Wavelet-based Image Coding Over Packet Networks", *Journal of Networks*, Vol 7, No 7 (2012), pp. 1046-1053, Jul 2012 doi: 10.4304/jnw. 7. 7. 1046-1053.
- [3] Jun Lu, Tong Wang, and Daxin Liu, "Research on Ordinal Properties in Combinatorics Coding Method", *Journal of Computers*, Vol 6, No 1 (2011), 51-58, Jan 2011 doi: 10.4304/jcp. 6. 1. 51-58.
- [4] Dawei Ding, Jun Yang, Qing Li, Liping Wang, and Wenyin Liu, "Towards a Flash Search Engine Based on Expressive Semantics", *Proceeding of ACM Multimedia 2003*, November 2003, Berkeley, CA.
- [5] Dawei Ding, Jun Yang, Qing Li, Liping Wang, and Wenyin Liu, "Automatic detection of Flash movie genre using Bayesian approach", *ICME 2004* pp. 603-606.
- [6] YANG Jun and LI Qing, "Content-based Retrieval of FlashTM Movies: Research issues, generic framework, and future directions", *Multimedia Tools and Applications*, 2007(34) pp. 1-23.
- [7] Adobe Corporation, Flash File Format Specification. <http://www.adobe.com/devnet.html>.
- [8] Meng Xiangzeng, "Structure analysis of Flash network teaching resource", *E-education Research*, 2010 (10) pp. 81-85.
- [9] Swfmill. <http://swfmill.org/>.
- [10] JavaSWF. <http://www.anotherbigidea.com/javaswf/>.
- [11] Liu Fei, Yu Wenchao. The Extraction of Tag Information in Flash Animation. *Computer Knowledge and Technology*, 2009(16) pp. 4277-4278.
- [12] Md. Mahbubur Rahman, "Unsupervised Natural Image Segmentation Using Mean Histogram Features", *Journal of Multimedia*, Vol 7, No 5 (2012), pp. 332-340, Oct 2012 doi: 10.4304/jmm. 7. 5. 332-340.
- [13] Mengzhao Yang, Kuanquan Wang, and Lei Zhang, "Realistic Real-time Facial Expressions Animation via 3D Morphing Target", *Journal of Software*, Vol 8, No 2 (2013), 418-425, Feb 2013 doi: 10.4304/jsw. 8. 2. 418-425.
- [14] Declan Kieran, Jonathan Weir, WeiQi Yan, "A Framework For An Event Driven Video Surveillance System", *Journal of Multimedia*, Vol 6, No 1 (2011), pp. 3-13, Feb 2011 doi:10.4304/jmm.6.1.3-13.



**ZHU Xiao-wei** is a teacher at Harbin University of Science & Technology. Her current main research fields are Digital Media, Computer Art and Design, and Computer Animation Design.