Abstract—Along with the ever-growing Web, horror contents sharing in the Internet has interfered with our daily life and affected our, especially children’s, health. Therefore horror image recognition is becoming more important for web objectionable content filtering. This paper presents a novel context-aware multi-instance learning (CMIL) model for this task. This work is distinguished by three key contributions. Firstly, the traditional multi-instance learning is extended to context-aware multi-instance learning model through integrating an undirected graph in each bag that represents contextual relationships among instances. Secondly, by introducing a novel energy function, a heuristic optimization algorithm based on Fuzzy Support Vector Machine (FSVM) is given out to find the optimal classifier on CMIL. Finally, the CMIL is applied to recognize horror images. Experimental results on an image set collected from the Internet show that the proposed method is effective on horror image recognition.

Keywords—Horror Image Recognition; Context-aware Multi-Instance Learning; Image Emotion

I. INTRODUCTION

Since the beginning of the Internet age, a vast amount of information, such as text, images, and videos, can now be shared from continent to continent, from country to country. Meanwhile, some materials, including pornography, horror and violence etc., can also easily interfere with our daily life and affect our, especially children’s, health. To protect our psychological health, lots of scientific researchers investigate into web filtering and try to filter out the objectionable contents automatically. Some pornographic or violent content filters, have matured to a point where robust recognition or filter software is available [1][2]. In contrast, the research of affective semantics of horror images is still on the stage of exploration. Lately, many psychological and physiological researchers emphasize the severity of horror images, primarily to children [3].

The early work on horror image recognition can be dated back to a part of affective semantics analysis [4] whose final goal is to invest the relationship between human affects and low level image features based on multiple emotions’ classification. And the horror image can be picked up if a type of fear emotion is recognized. For example, Yanulevskaya et al [5] classify images into 10 emotional categories from holistic texture features. Solli et al [6] propose a color based Bag-of-Emotions model for image retrieval from emotional viewpoints. Datta et al[18] analyze the image’s emotional semantic from aesthetic aspect. Essence in all of these methods [4][5][6][17][18] is to describe holistic low level visual feature and use machine learning techniques to bridge the gap between these features and the interpretation of different emotions. Recently, many researchers have begun to investigate horror image recognition as a specific topic. A new aspect is region-based method proposed by Li et al [7] that uses local regions and specific characteristics embedded in horror images. The potential problem in this method[7] is that some non-horror images may be misclassified since extraction and analysis of all the regions are processed independently.

In this paper, we present a novel horror image recognition model based on context-aware multi-instance learning (CMIL). The model assumes that, for any horror image, there may be no definitely isolated horror region in it, but exists at least one region that can express horror feeling through contextual cooperation. The proposed method segments the image into regions and characterizes each region by a group of feature vectors describing color, texture, and emotion attributes. Consequently, an image can be represented by a collection of feature vectors as well as their contextual relations through an undirected graph. Then the CMIL is used to classify the horror images considering the contextual information among regions. The method that is close in spirit to ours is miGraph and MiGraph proposed by Zhou et al [9]. The main difference is that Zhou et al [9] consider the global contextual structural relationship, whereas we focus on local interactions.

II. CONTEXT-AWARE MULTI-INSTANCE LEARNING

A. MIL and Horror Image Recognition

In this section, we firstly review the traditional MIL. Let \( \chi \) denote the instance space. Given a data set \( \{(X_1, Y_1), ..., (X_n, Y_n)\} \), where \( X_i = \{x_{i,1}, ..., x_{i,j}, ..., x_{i,n}\} \subseteq \chi \) is called a bag and \( x_{i,j} \in \chi \) is an instance. \( Y_i \in \Psi = \{-1, +1\} \) is the label of bag \( X_i \), and \( y_{i,j} \in \Psi \) is the underlying label of instance \( x_{i,j} \). In MIL, these labels are interpreted in the following way: if \( Y_i = -1 \), then \( y_{i,j} = -1 \) for all instances in bag \( X_i \), i.e.
no instance in the bag is positive. On the other hand, if \( Y_i = +1 \), then at least one instance \( x_{i,j} \in X_i \) is positive, i.e. \( y_{i,j} = +1 \). Notice that the information provided by the bag label is asymmetric in the sense that a negative bag label induces a unique label for every instance in a bag; while a positive label does not. So far, many computational models have been proposed for MIL [8][10][9].

MIL may be employed in horror image recognition; because it has been shown that each horror image always includes at least one horror region [7]. The potential solution is to acquire objects in the image and feed them into MIL for recognition. However, the emotion that the same object evokes depends on not only visual scene itself but also the environment around it. As a result, MIL may inevitably fail to identify the two images’ emotions only based on each object’s independent contexts. The reason for this is that their environments are different. But MIL inevitably fails to identify the two images’ emotions only based on each object’s independent features.

To address the problem, we extend MIL and build up a Context-Aware Multi-Instance Learning Model (CMIL) by adding contextual correlations between instances. The objective of the model is to properly formulate the horror image recognition problem, because there may be no definitely positive isolated instance in some positive bags; but there is at least one instance definitely expressing positive property through contextual cooperation.

B. Formulation of CMIL

Now the data set for the CMIL model is redefined as: \( \{(X_1,M_1,Y_1),...,(X_i,M_i,Y_i),...,(X_N,M_N,Y_N)\} \), where \( X_i \), \( x_{i,j} \) and \( Y_i \) are a bag, an instance and the label respectively, the same as those definitions in MIL. The only difference is a new added item \( M_i \), which is a contextual graph that represents the relationship among different instances in the bag \( X_i \). Its structure can be from practical applications’ requirements. Also different from MIL, the underlying label of instance \( x_{i,j} \) in CMIL is a fuzzy label, defined as \( < y_{i,j},s_{i,j} > \in \theta_i \), where \( y_{i,j} \in \Psi = \{-1,+1\} \) is the class label of \( x_{i,j} \); and \( 0 < s_{i,j} \leq 1 \) is the fuzzy membership associated with instance \( x_{i,j} \); and \( \theta_i \) is the label and fuzzy membership set of the bag \( X_i \). A contextual energy function \( E_{i,j}(\theta_i,M_i) \) is also given for each instance \( x_{i,j} \) in the bag \( X_i \) under the contextual graph \( M_i \). The function can be regarded as the attitude of the corresponding instance toward the positive class including contextual information. Consequently, the labels of bags in the CMIL model are interpreted as: If \( Y_i = +1 \), then at least one instance \( x_{i,j} \in X_i \) has \( E_{i,j}(\theta_i,M_i) \geq 1 \). If \( Y_i = -1 \), then \( E_{i,j}(\theta_i,M_i) < 1 \) for all the instances in bag \( X_i \).

Comparing with MIL, we can find that our CMIL model makes differences in the following aspects: (1) Input bag not only contains instance’s data and bag’s label, but also contains a contextual graph that represents the contextual relationships among different instances in the same bag. (2) The underlying label of the instance in CMIL is a fuzzy label. (3) The label of the bag is not decided by instances’ labels themselves, but depends on the contextual energy of each bag.

1) Contextual Map: Here, we define the contextual graph \( M_i \) as an undirected graph. An example is shown in Figure 2(A). Each vertex of \( M_i \) is an instance; and there is an edge if there is a direct contextual link between two instances. Moreover, assume that the length of each edge in \( M_i \) is always 1 unit. If there exists a shortest path from \( x_{i,j} \) to \( x_{i,k} \) with \( n \) units, we say that \( x_{i,k} \) is \( n \)-contextual-neighbor (denoted as \( CN^n(x_{i,j}) \)) of \( x_{i,j} \). All the \( n \)-contextual-neighbors compose a set for each instance, as \( CN^1(x_{i,j}) = \{CN^1(x_{1,j}),CN^2(x_{i,j}),...\} \). The \( n \)-contextual-neighbors can be found by the Breadth-first search (BFS) tree in graph \( M_i \). Figure 2 (B) shows that all the \( n \)-contextual-neighbors of the instance \( x_{i,1} \).

2) Energy Function: Based on the contextual graph and the membership set, we define an energy function \( E_{i,j}(\theta_i,M_i) \), for each instance \( x_{i,j} \) in the bag \( X_i \). The
function is the core of the CMIL model and should reflect the degree of each instance belonging to the positive label with contextual information. The larger the $E_{i,j}(\theta_i, M_i)$ is, the more possibly positive the $x_{i,j}$ is. Therefore, the function defined in the paper includes two parts:

$$E_{i,j}(\theta_i, M_i) = y_{i,j} \times s_{i,j} + E^*(CN(x_{i,j})), \quad (1)$$

where $y_{i,j} \times s_{i,j}$ is the independent energy of $x_{i,j}$, which expresses the class label of the instance $x_{i,j}$; while $E^*(CN(x_{i,j}))$ represents the energy contribution from its contextual neighbors. From the principle that the instance near to $x_{i,j}$ should contribute more energy, we further define $E^*(CN(x_{i,j}))$ as:

$$E^*(CN(x_{i,j})) = \max_{1 \leq p \leq m} \left\{ \max_{x_{i,k} \in CN^p(x_{i,j})} \left(2^{-p} \times y_{i,k} \times s_{i,k}\right) \right\}, \quad (2)$$

where $m$ is a distance parameter showing that only the contextual neighbors in $CN^1(x_{i,j}) \cup CN^2(x_{i,j}) \cup ... \cup CN^m(x_{i,j})$ are considered. The factor of $(1/2)^p$ makes sure that the farther the neighbor, the weaker is its effect on $x_{i,j}$.

Therefore, the label of each bag in CMIL can be described as a constraint:

$$Y_i \times \left( \max_{1 \leq j \leq n_i} (E_{i,j}(\theta_i, M_i)) - 1 \right) \geq 0. \quad (3)$$

C. Optimize Classifier on CMIL via Fuzzy SVM

Inspired by Support Vector Machines (SVM) for MIL [10] and noticing that a fuzzy system is introduced in the CMIL model, we extend the Fuzzy Support Vector Machine (FSVM) [11] to optimize the classifier in the CMIL model.

1) Maximum Pattern Margin via FSVM: The SVM has been extended as mi-SVM and MI-SVM to solve MIL problems [10][12]. However, the labels for the training samples in these methods are binary: each one belongs to either one class or the other. By introducing fuzzy theory, Lin et al [11] propose a FSVM. Because of space limitation, we skip the details of SVM, mi-SVM as well as FSVM; and refer interesting readers to excellent references [10][11]. In this paper, we propose CMIL-FSVM by adding the energy function as another constraint in FSVM.

Assuming that the classification hyperplane is $f(x) = w^T x + b$, inspired by FSVM’s optimization object, the optimization object function of CMIL can be rewritten as:

$$\min_{w} \frac{1}{2} \|w\|^2 + C \sum_{i} \sum_{j=1}^{n_i} s_{i,j} \xi_{i,j}, \quad (4)$$

subject to $y_{i,j} (w \cdot x_{i,j} + b) \geq 1 - \xi_{i,j}, \xi_{i,j} \geq 0, \quad (5)$

$$Y_i \times \left( \max_{1 \leq j \leq n_i} (E_{i,j}(\theta_i, M_i)) - 1 \right) \geq 0. \quad (6)$$

Notice that, in the standard classification of FSVM, the labels $y_{i,j}$ and fuzzy membership $s_{i,j}$ are predefined; instead both of them will only be given out implicitly in the CMIL-FSVM.

2) Optimization Heuristics: The same as the analysis in [10], CMIL-FSVM can be viewed as a mixed-integer problem. Therefore, given instances’ hidden labels and fuzzy memberships, it can be reduced to a Quadratic Programming (QP) and solved exactly through FSVM.

Borrowing the strengths of optimization procedure in mi-SVM, we further present a simple optimization heuristic algorithm based on FSVM. It employs an iterative computation to minimize the object function (Equation 4) with a discriminant function constraint (Equation 5) and the energy function constraint (Equation 6).

During the initialization stage, we pair positive labels of those instances in positive bags with low fuzzy memberships, and negative labels of those instances in negative bags with higher fuzzy memberships so as to make sure that all the initial energy functions for negative bags are less than 0. Then, we iteratively adjust the hyperplane by improving the memberships to let positive bags satisfy the constraint defined in Equation (6) until convergence and

<table>
<thead>
<tr>
<th>Table I</th>
<th>PSEUDO-CODE FOR CMIL-FSVM OPTIMIZATION HEURISTICS.</th>
</tr>
</thead>
</table>
| **Training Procedure:** | Initialize $y_{i,j} = Y_i$, for $x_{i,j} \in X_i$, $s_{i,j} = 1$, for $x_{i,j} \in X_i$ and $Y_i = -1$, $s_{i,j} = 0.5$, for $x_{i,j} \in X_i$ and $Y_i = +1$.
Set iteration times $IT = 0$, $\Delta s = 0.1$, $threT = 100$.

REPEAT
Set $IT = IT + 1$.
Compute FSVM solution for data set with imputed labels.
Compute outputs $f_{i,j} = w^T x_{i,j} + b$ for all instances.
set $y_{i,j} = sgn(f_{i,j})$, $s_{i,j} = FS(f_{i,j})$ for all $x_{i,j}$ all in positive bags.
Set $p = 0$.
FOR (every positive bag $X_i$)
Compute $j^* = \arg \max_{1 \leq j \leq n_i} (E_{i,j}(\theta_i, M_i))$;
IF $(E_{i,j}(\theta_i, M_i) < 1)$
Set $s_{i,j} = \min \left(1, \frac{s_{i,j}}{1 + \delta_{i,j}^p} \right)$;
Set $p = p + 1$;
END
END
WHILE($\Delta s = 0$ OR $IT < threT$).
OUTPUT($w$, $b$).

| **Test Procedure:** | Compute outputs $f_{i,j} = w^T x_{i,j} + b$ for all $x_{i,j}$ in $X_k$.
Set $y_{k,j} = sgn(f_{k,j})$, $s_{k,j} = FS(f_{k,j})$ for all $x_{k,j}$ in $X_k$.
Set $j^* = \arg \max_{1 \leq j \leq n_k} (E_{k,j}(\theta_k, M_k))$.
IF $(E_{k,j}(\theta_k, M_k) \geq 1)$
Set $Y_k = 1$;
ELSE
Set $Y_k = -1$;
END. |
then produce the final classifier. So the iterative procedure is involved into two major steps: (1) For given instances’ labels and fuzzy memberships, solve the associated QP and find the optimal discriminate function. (2) For the calculated labels and fuzzy memberships, solve the associated QP and is involved into two major steps: (1) For given instances’ labels and fuzzy memberships, solve the associated QP and find the optimal discriminate function, update partial or all instances’ labels and memberships in a proper way that locally minimizes the objective function. The implementation details are shown in Table 1. The function $FS(t)$ in Table 1 is defined as:

$$FS(t) = \begin{cases} 1, & \text{if } |t| > 1, \\ |t|, & \text{else.} \end{cases}$$

(7)

III. HORROR IMAGE RECOGNITION BASED ON CMIL

Horrors image recognition is actually a typical CMIL problem because the horror regions that can evoke fear emotion may be not horrific independently, but play a role through contextual cooperation with others.

A. Contextual Map Generation

The first step in horror image recognition based on CMIL is to construct instances, bags and contextual maps. The intuitive idea here is to view the whole image as a bag and segmented regions in the image as instances. Among diverse segmentation algorithms, JSEG algorithm [13] is adopted for its flexibility of adjusting the number of regions. After segmentation, we remove the regions whose areas are smaller than 1/20 of the whole image. Figure 3(B) shows an example result of JSEG.

The following stage is to construct contextual graph using these regions from the image. Context knowledge is recently widely used for visual categorization and image understanding [14]. It can be roughly divided as semantic context, spatial context, scale context etc [14]. The spatial context, a kind of direct contextual knowledge, is adopted in the form of an adjacency graph in this paper. Therefore, the vertexes of the graph represent those selected regions in the image and an edge with length of 1 is defined between any two adjacent regions. The example contextual graph of an image in Figure 3(A) is shown in Figure 3(C).

B. Feature Extraction

In this section, three types of features, color, color emotion, as well as color texture, for each region are extracted; because algorithms in [6][5][7] have proved their effectiveness in horror image recognition.

Color Feature. We consider the image color information in HSV space. The averaged HSV color values of each region and the averaged HSV color values of the remaining pixels of the input image are concatenated into a 6-dimension color feature vector.

Color Emotion Feature. Color emotion refers to the emotion caused by a single color and has long been of interest to both artists and scientists. Recent research of Ou et al [15] gives a 3D computational color emotion model, each dimension representing color activity (CA), color weight (CW), and color heat (CH) respectively. The transformation equations between the color space and the color emotion space are defined as:

$$CA = -2.1 + 0.06 \left[ (L^* - 50)^2 + (a^* - 3)^2 + \left( \frac{b^* - 17}{1.4} \right)^2 \right]^{1/2}$$

$$CW = -1.8 + 0.04(100 - L^*) + 0.45 \cos(h^* - 100\circ)$$

$$CH = -0.5 + 0.02(C^*)^{1.07} \cos(h^* - 50\circ)$$

(8)

where $(L^*, a^*, b^*)$ and $(L^*, C^*, h^*)$ are the color values in the CIELAB and CIELCH color spaces respectively. We also calculate the averaged color emotion values of each region and the averaged color emotion values of the remaining pixels in the image and compose them as another 6-dimension color emotion feature vector.

Texture Feature. Geusebroek et al [16] report a six-stimulus basis for stochastic texture perception. Fragmentation of a scene by a chaotic process causes the spatial scene statistics to conform to a Weibull-distribution as:

$$wb(x) = \frac{\gamma}{\beta} \left( \frac{x}{\beta} \right)^{\gamma - 1} e^{-\left( \frac{x}{\beta} \right)^{\gamma}},$$

(9)

where the parameters of the distribution can completely characterize the spatial structure of the texture. For example, the contrast of an image can be represented by the width of the distribution $\beta$, and the grain size is given by $\gamma$, which is the peakedness of the distribution.

The Weibull texture features for an image region consist of the Weibull parameters for the color edges in the region. Thus, the $\beta$ and $\gamma$ values for the $x$-edges and $y$-edges of the three color channels yields a 12 dimensional feature vector for texture.

Therefore, if we put these three types of features together, we can represent each segmented region with a 24-dimensional vector.

C. Horror Image Recognition Based on CMIL

After getting the contextual graph of the image and feature values of each image region, we can construct a bag for CMIL. The bag $X_i$ is the whole image; the instance $x_{i,j}$ is the feature vector of each region in the image; $M_i$ is the spatial contextual graph for the image; and the label $Y_i$ for each image is set as 1 if it is a horror image, set as -1
otherwise. Then the CMIL-FSVM is applied to classify the horror images.

IV. EXPERIMENTS

To evaluate the performance of the proposed scheme, we conduct experiments and compare it with other methods on an image set collected from the Internet. The competitors include Horror Image Recognition based on Emotional Attention (HEA) [7], the Emotional Valence Categorization (EVC) algorithm [5] and the color based Bag-of-Emotions (BoE) model [6]. Although the EVC and BoE are not specially designed for horror image recognition, we can still use them to classify horror and non-horror emotions. In order to validate the effectiveness of the proposed CMIL for horror image recognition, two other popular multi-instance learning algorithms, mi-SVM and MI-SVM [10], as well as two classifiers with contextual information, miGraph and MIGraph [9], are also considered.

We divide the algorithms included in this paper into 3 categories, Global, Local and Contextual, based on the primary information they use for recognition. The EVC and BoE, which use the features from the entire image to recognize emotions, belong to the global group. The MI-SVM and mi-SVM are the local methods, because they only consider regions’ feature for horror image recognition. The contextual methods are HEA, miGraph, MIGraph and CMIL algorithms because they consider relationship between segmented regions.

To make sure that all of these methods are comparable, the bag construction and instance’s features used in these methods are the same as those used in our method. Moreover, the Radial Basis Function (RBF) is adopted as the kernel function in CMIL-FSVM, mi-SVM, MI-SVM, miGraph and MIGraph. The optimized parameters for kernel functions for each algorithm are determined through applying 3-cross-validation on the training set in each experiment. In order to simplify computation, we set $m = 1$ in Equation (2) for the proposed algorithm in the following experiments.

A. Data Set and Error Measurement

Because of lack of open large data sets for horror image recognition, we collected and created a horror image set from the Internet. The images are from three favorite search engines: google.com, bing.com and baidu.com.

This horror image set includes 500 horror images and 500 non-horror images. A large number of candidate horror images are firstly collected from three image search engines (google.com, bing.com, baidu.com). Then seven students in our Lab are asked to label them from one of the three categories: Non-horror, A little horror, and Horror. Then 500 horror images are picked out, each of which is labeled as ‘Horror’ by at least 4 users. On the other hand, we also collect 500 non-horror images with different scenes, objects or emotions. Specially, the non-horror images includes indoor images, 50 outdoor images, 50 human images, 50 animal images, 50 plant images, and 250 images with different emotions (adorable, amusing, boring, exciting, irritating, pleasing, scary, and surprising) that are downloaded from an image retrieval system ALIPR(http://alipr.com/) with emotional query words.

For the horror image set, given the ground truth of a horror image set as $H_S$, and the recognition results $E_S$ from an algorithm, the precision ($pre$), recall($rec$), and $F_1$ measure defined in Equation (10) are used to evaluate the performances of diverse algorithms.

$$pre = \frac{|HS \cap ES|}{|ES|}, \quad rec = \frac{|HS \cap ES|}{|HS|}, \quad F_1 = \frac{2 \times prec \times rec}{prec + rec}$$

B. Experimental Results

We equally divide the image set into two subsets A and B. Each subset contains 250 horror and 250 non-horror images. We then use A for training and B for test and vice versa. The combined results of the two experiments are used as the final performance. The experiment results shown in Table 2.

The local algorithms (mi-SVM, MI-SVM) and contextual algorithms (HEA, miGraph, MIGraph and CMIL) are slightly better than global ones. This phenomenon invalidates our previous statement that horror emotion is most likely evoked by local regions. Using holistic information from the image misclassifies some horror images with large area background. For example, the two images in Figure 4(A) are difficult to be classified correctly only from global information.

<table>
<thead>
<tr>
<th>Method</th>
<th>$pre$</th>
<th>$rec$</th>
<th>$F_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVC</td>
<td>0.760</td>
<td>0.622</td>
<td>0.684</td>
</tr>
<tr>
<td>BoE</td>
<td>0.741</td>
<td>0.600</td>
<td>0.663</td>
</tr>
<tr>
<td>mi-SVM</td>
<td>0.702</td>
<td>0.722</td>
<td>0.712</td>
</tr>
<tr>
<td>MI-SVM</td>
<td>0.684</td>
<td>0.706</td>
<td>0.695</td>
</tr>
<tr>
<td>HEA</td>
<td>0.768</td>
<td>0.790</td>
<td>0.769</td>
</tr>
<tr>
<td>mi-Graph</td>
<td>0.723</td>
<td>0.748</td>
<td>0.735</td>
</tr>
<tr>
<td>MI-Graph</td>
<td>0.721</td>
<td>0.715</td>
<td>0.718</td>
</tr>
<tr>
<td>CMIL</td>
<td>0.792</td>
<td>0.776</td>
<td>0.784</td>
</tr>
</tbody>
</table>

Figure 4. Examples of some easy misclassification images.
Among all the local algorithms, mi-SVM and MI-SVM, which consider image regions independently, have lower performances than the contextual methods HEA and CMIL. The local algorithms mi-SVM and MI-SVM may avoid the misclassification for images like Figure 4(A), but it is still difficult for them to differentiate two images in Figure 4(B). It is due to the fact that the regions with the girl are so very similar if we only consider local features. It also implies that horror image recognition can benefit from exploiting contextual information among the image’s regions.

As a matter of fact, our proposed CMIL algorithm outperforms all other methods, with \( pre = 0.792, \) \( rec = 0.776 \) and \( F_1 = 0.784 \). The HEA algorithm, which achieves good performance in this image set, also takes the contextual information into account. But the contextual information used in HEA is the interaction between emotional salient regions and average features of the whole images; therefore it depends on the performance of saliency map pre-computation. In addition, the average of all features embedded in the whole images sometimes cannot express contextual information about horror emotion correctly. Although the miGraph and MIGraph also take the contextual information into account, they consider all regions based on Euclidean distances among them and focus on global graph structures. By comparison, our method models the spatial contextual information between regions in the recognition. So it can selectively pay more attention to those most possible contextual regions that can promote the image as a horror image. Therefore, our algorithm can classify images in Figure 4(A) and (B). However, we also noticed that the proposed algorithm fails for the images shown in Figure 4(C), because the contextual information embedded in them is very similar to those from horror images.

V. CONCLUSION

In this paper, we propose a context-aware multi-instance learning (CMIL) model to automatically recognize horror images. Our work is based on the fact that the horror emotion is not evoked by some independent regions in the images but by the relations between them. A modified FSVM with the additional energy function’s constraint is also proposed to optimize the CMIL classifier. To prove effectiveness of the proposed CMIL model, we evaluate our system on a real image set. Experimental results show that our proposed algorithm is superior to the others.

ACKNOWLEDGMENT

This work is supported by National Nature Science Foundation of China (No. 61005030, 60935002 and 60825204) and the Excellent SKL Project of NSFC (No.60723005).

REFERENCES