SEMANTIC DETECTION OF ADULT IMAGE USING SEMANTIC FEATURES

Jae-Hyun Jeon, Se Min Kim, Jae-Young Choi, Hyun Suk Min, and Yong-Man Ro

Image and Video Systems Lab, Department of Electrical Engineering,
Korea Advanced Institute of Science and Technology (KAIST), Yuseong-gu, Daejeon, 305-732, Korea

ABSTRACT

Recently, in the fields of internet and social networking, the classification and filtering of naked images has been receiving a significant amount of attention. In this paper, we propose a novel naked image classification which can make effective use of semantic features of a naked image. In addition, a novel measurement, termed accumulated distance ratio (ADR), is proposed in order to systematically analyze the effect of semantic features on improving classification performance, compared to the approach relying on low-level visual features. Extensive experiments have been carried out to assess the effectiveness of semantic features in naked image classification with realistic and challenging data set. The experimental result of the proposed approach using semantic features, for challenging data set, shows improvement up to 14% than the approach using low-level visual feature. Further, the proposed ADR measure has proven to be useful measure for analyzing the effect of semantic features for naked image classification.

Index Terms— Naked image classification, Semantic features, Low-level visual features, Accumulated distance ratio, SVM

1. INTRODUCTION

With the rapid growth of the Internet, any user can access and browse multimedia contents posted on the Web. Although the Internet brings convenience, adult contents (such as naked images) are widely available. These adult contents may be harmful to users, particularly to the children and adolescence. As such, classifying and filtering the adult contents has been being popular in the Internet.

Thus far, considerable research efforts have been dedicated to naked image classification techniques. Most previously developed naked image classification techniques have been limited to using low-level visual feature including color, edge, and texture features. In order to classify naked images from non-naked images, low-level visual feature is known to be enough to obtain reliable classification accuracy [1-2]. However, one weakness of previous approaches making use of only low-level visual features is that non-naked images *containing bikinied woman* are not well classified from naked images, because low-level visual features of bikinied images are not much different from naked images. Namely, false positive and false negative images in classification results are increased due to bikinied images. For example, in [1], it is shown that bikinied images are worse classified than other non-naked images. In [2], when bikinied images are added into non-naked images, the classification performance is decreased. Thus, the classification of naked images from bikinied images is important in naked image classification.

Semantic feature has been successfully used in the areas of scene image classification [5-6]. The semantic feature based classification technique has used the information of semantic concepts existing in image [5]. Hence, we believe that semantic concepts (detected from adult image) could be effective with high discriminatory power for the purpose of classifying naked and bikinied images. However, there are few attempts to make use of semantic feature and to analyze the effect of semantic feature compared with low-level visual feature for naked image classification [4]. In this paper, we present useful semantic features suitable for naked image classification. In particular, we devise effective semantic features in terms of classifying naked images from bikinied images. For classification purpose, we incorporate the associated semantic feature into Support Vector Machine (SVM) classifier [5]. Through extensive experiments, we demonstrate that the proposed semantic feature is more useful for classifying naked image and more robust to the novel test image unseen in training stage than low-level visual feature. Furthermore, we derive an accumulated distance ratio (ADR) measurement in SVM in order to make a thorough analysis of the effect of semantic feature in improving classification performance, compared with low-level visual feature.

The rest of the paper is organized as follows: Section 2 outlines the naked image classification framework using semantic feature. Section 3 presents the analysis for the effect of semantic features using ADR. Experimental results are presented in Section 4. Finally, conclusion is drawn in Section 5.

2. SEAMANTIC FEATURE BASED NAKED IMAGE CLASSIFICATION USING SVM

In this section, we discuss the proposed naked image classification based on semantic features. The overall framework of our naked image classification method is shown in Fig. 1. Note that the proposed framework is largely composed of two parts: 1) the generation of semantic features and 2) the classification of the naked images. The details of these two parts will be explained in the following subsections.

2.1. Generating semantic features to classify naked images

The goal of this section is to describe the generation of semantic features. Up to the present, there seems no almighty method for object segmentation. Most related approaches are quite expensive in computation and even sometimes produce incomplete result in complex images. So, instead, we use a simple block segmentation

978-1-4244-7566-7/10/$26.00 ©2010 IEEE

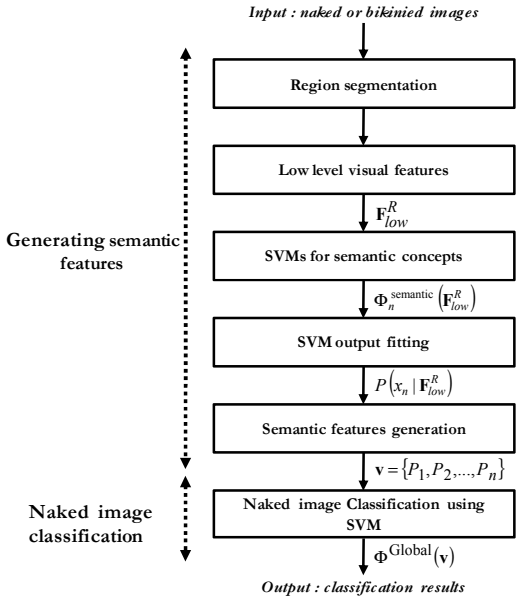

Fig. 1. Overall framework of the proposed naked image classification based on semantic features.

to capture semantic concepts that appear on local photo regions. In this paper we employ the regional segmentation approach proposed in [5]. If either a naked or bikini image (to be classified) is first entered to classification system, an input image is divided into the m regions in which adult semantic concepts are frequently observed in images. Then, low-level visual features are extracted from the associated segmented region in the following way:

$$\mathbf{F}_{low}^{R}=\left[\left(\mathbf{f}_{low}^{1}\right)^{\mathrm{T}}\left(\mathbf{f}_{low}^{2}\right)^{\mathrm{T}}\cdots\left(\mathbf{f}_{low}^{L}\right)^{\mathrm{T}}\right]^{\mathrm{T}}, \tag{1}$$

where $\mathbf{f}_{low}^{(i)}\left(1\le i\le L\right)$ can be generalized to including color, texture,and edge-related features (or descriptors), L is the number of low-level visual features considered, and T stands for the transpose operator of a matrix. We employ MPEG-7 visual features for the low-level visual features.

In the process of defining semantic concepts for naked images, there are three key characteristics taken into consideration: 1) feasibility, 2) observability, and 3) utility [6-7]. *Feasibility* represents capability of how well a certain adult semantic concept can be represented by a predefined set of low-level visual features. Further, *observability* indicates that one has to select the adult semantic concepts (of all possibly used semantic concepts) that occur with a high frequency within a given adult image set. Lastly, *utility* indicates that the selected semantic concepts should be useful for the target naked image classification.

We devise the semantic concepts (based on the aforementioned three characteristics) that are used in our classification system. Let $\mathbf{X}=\{x_1,x_2,...x_N\}$ be a set consisting of N semantic concepts (x_n) which are defined in the training stage such as 'body', 'breast', 'genital', 'bottom', 'bikini', 'brassiere', 'panty'.

Individual binary SVM classifier is separately and independently trained with each corresponding semantic concept for classification purpose. Let us denote that $\Phi_n^{\text{semantic}}\left(\mathbf{F}_{low}^{R}\right)$ is the output of the n^{th} semantic SVM (trained with the corresponding n^{th} semantic concept x_n), obtained from the associated segmented region (R). The $\Phi_n^{\text{semantic}}\left(\mathbf{F}_{low}^{R}\right)$ is computed as follows:

$$\Phi_n^{\text{semantic}}\left(\mathbf{F}_{low}^{R}\right)=\sum_t\left\{\mathbf{w}_n(t)\cdot\mathbf{z}_n(t)\cdot K\left(\mathbf{g}_n(t),\mathbf{F}_{low}^{R}\right)\right\}+a_n, \tag{2}$$

where K is kernel function, $\mathbf{g}_n(t)$ is the t^{th} support vector of the hyper-plane for x_n, $\mathbf{w}_n$ is the corresponding weighting vector of the support vector, $\mathbf{z}_n$ is the corresponding class vector of the support vector, and a_n is the threshold value optimized for x_n.

Standard SVMs do not provide such probabilities. Thus, each output of a corresponding semantic SVM should be a calibrated posterior probability to enable post-processing. Then, each output $\Phi_n^{\text{semantic}}\left(\mathbf{F}_{low}^{R}\right)$ of a corresponding semantic SVM is fitted to a parametric sigmoid model as follows:

$$P\left(x_n \mid \mathbf{F}_{low}^{R}\right)\cong\frac{1}{1+\exp\left(A_n\cdot\Phi_n^{\text{semantic}}\left(\mathbf{F}_{low}^{R}\right)+B_n\right)}, \tag{3}$$

where A_n and B_n are parameters used to adjust the shape of the sigmoid model for the semantic concept (x_n) and measured by solving the regularized maximum likelihood problem [8]. Thus the SVM output ranging from -∞ to ∞ should be mapped to the probabilistic output ranging from 0 to 1.

In order to obtain the semantic feature, each probability of a corresponding semantic concept (x_n) for an input image should be computed from probabilities of the corresponding semantic concept (x_n) obtained from the segmented regions in advance. Then aforementioned probability for an input image is computed as follows:

$$P_n=\max\left\{P\left(x_n \mid \mathbf{F}_{low}^{R_1}\right),P\left(x_n \mid \mathbf{F}_{low}^{R_2}\right),...,P\left(x_n \mid \mathbf{F}_{low}^{R_m}\right)\right\} \tag{4}$$

Note that P_n represents the probability that a certain image contains the semantic concept x_n. Therefore, the semantic feature of the image should be obtained from n probabilities that the image contains n corresponding semantic concepts as follows:

$$\mathbf{v}=\{P_1,P_2,...,P_n\}. \tag{5}$$

2.2. Naked image classification

Semantic features of the image are classified with global SVM (trained to classify the semantic features of naked images). The output of global SVM is expressed as shown in (6). Finally, If $\Phi^{\text{Global}}(\mathbf{v})>0$, input image is classified for the 'naked' image.

$$\begin{aligned}\Phi^{\text{Global}}(\mathbf{v})&=\sum_t\left\{\mathbf{w}(t)\cdot\mathbf{z}(t)\cdot K\left(\mathbf{g}(t),\mathbf{v}\right)\right\}+a\\&=\mathbf{w}^{\mathrm{T}}\mathbf{v}+a,\end{aligned} \tag{6}$$

where K is kernel function, $\mathbf{g}(t)$ is the t^{th} support vector of the hyper-plane for 'naked' concept, $\mathbf{w}$ is the corresponding

weighting vector of the support vector, **z** is the corresponding class vector of the support vector, and a is the threshold value optimized for 'naked' concept.

3. EVALUATION OF USED FEATURES USING ADR

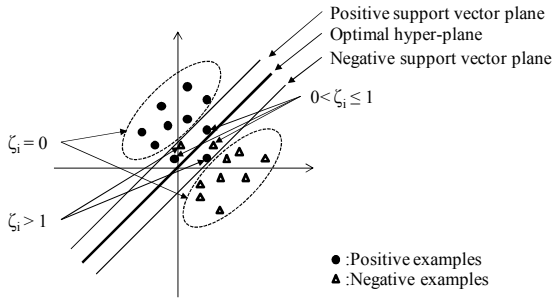

Fig. 2. The distribution of positive and negative training samples in the hyper-plane of SVM. Note $\zeta_i = \zeta_j^+ \text{ or } \zeta_k^-$.

As discussed in Section 2, in our framework, the semantic features are first generated and then naked images are classified using SVM with the semantic features. The scatter-matrix-based measure is often used to evaluate the used features [9]. But, this measure is designed without consideration of hyper-plane of SVM. To evaluate the effect of used features related to hyper-plane of SVM, we propose accumulated distance ratio (ADR). This is mainly attributed to the fact that the classification performance of SVM depend highly upon the positions of training samples with respect to the hyper-plane of SVM.

3.1. Hyper-plane in SVM

Fig. 2 shows an example of the distribution of positive and negative training samples in hyper-plane of SVM. In the hyper-plane of SVM, optimal hyper-plane and two support vector planes can be expressed as follows:

$$\begin{aligned}
&\text{Optimal hyper - plane}: \mathbf{g}(\mathbf{v}) = \mathbf{w}^{\mathrm{T}}\mathbf{v} + a = 0,\\
&\text{Positive support vector plane}: \mathbf{g}^{+}(\mathbf{v}) = \mathbf{w}^{\mathrm{T}}\mathbf{v} + a = 1,\\
&\text{Negative support vector plane}: \mathbf{g}^{-}(\mathbf{v}) = \mathbf{w}^{\mathrm{T}}\mathbf{v} + a = -1,
\end{aligned} \tag{7}$$

where **v** is the feature vector to be used for SVM, **w** is the normal vector of the optimal hyper-plane.

Let us define positive and negative training samples as $\mathbf{v}_j^+$ and $\mathbf{v}_k^-$ ($1 \le j \le N^+, 1 \le k \le N^-$), where N^+ and N^- are the number of positive and negative training samples.

As shown in Fig. 2, we define ζ_j^+ and ζ_k^- as

$$\begin{aligned}
\zeta_j^+ &= \text{ramp}\left(1 - \mathbf{g}^{+}(\mathbf{v}_j^+)\right),\\
\zeta_k^- &= \text{ramp}\left(1 + \mathbf{g}^{-}(\mathbf{v}_k^-)\right),
\end{aligned} \tag{8}$$

where

$$\text{ramp}(v) = \begin{cases} v, v \ge 0 \\ 0, v < 0 \end{cases}. \tag{9}$$

3.2. Accumulated Distance Ratio to measure the effect of semantic features

Through the relation between values (ζ_j^+ and ζ_k^-) of training samples and classification performance of SVM, we found that the training samples having $\zeta_j^+, \zeta_k^- = 0$ meant correctly classified. On the other hand, the training samples having $\zeta_j^+, \zeta_k^- > 0$ meant possibly misclassified.

The effect of used features can be measured by accumulating distances between training samples and the support vector planes. Namely, the ratio between accumulated distances of training samples from positive and negative support planes can show the effect of used features. ADR can be written as

$$ADR = \frac{\sum_{j=1}^{N_{\text{zero}}^+} d\left(\mathbf{v}_j^+, \mathbf{g}^+\right) + \sum_{k=1}^{N_{zero}^-} d\left(\mathbf{v}_k^-, \mathbf{g}^-\right)}{\sum_{j=1}^{N_{\text{non-zero}}^+} d\left(\mathbf{v}_j^+, \mathbf{g}^+\right) + \sum_{k=1}^{N_{\text{non-zero}}^-} d\left(\mathbf{v}_k^-, \mathbf{g}^-\right)}, \tag{10}$$

where N_{zero}^+ and N_{zero}^- are the number of positive and negative training samples having $\zeta_j^+, \zeta_k^- = 0$, respectively while $N_{\text{non-zero}}^+$ and $N_{\text{non-zero}}^-$ are the number of positive and negative examples having $\zeta_j^+, \zeta_k^- > 0$ respectively. Note that $N_{\text{zero}}^+ + N_{\text{non-zero}}^+ = N^+$, $N_{\text{zero}}^- + N_{\text{non-zero}}^- = N^-$ and the distance between feature vector and plane is computed as follows:

$$d\left(\mathbf{s}_i, \mathbf{h}\right) = \frac{\mathbf{h}(\mathbf{s}_i)}{|\mathbf{u}|}, \tag{11}$$

where

$$\mathbf{h}(\mathbf{s}) = \mathbf{u}^{\mathrm{T}}\mathbf{s} + b \tag{12}$$

where $|\mathbf{u}|$ is a magnitude for the normal vector of a plane $\mathbf{h}(\mathbf{s})$.

Higher ADR value indicates that the number of training samples between support vector planes becomes sparse and simultaneously a lot of training samples having $\zeta_p^{(j)}, \zeta_n^{(k)} = 0$ are to be far from the support vector planes. In other words, from the point-of-view of test classification, it implies that the used features are more powerful to classify the unseen test samples.

4. EXPERIMENT

In order to verify the usefulness of the semantic features in the naked image classification, experiments were performed. In this section, we first describe our data sets and experiment conditions. Then, we report the experimental results with performance of naked image classification and measurement of ADR.

4.1. Data sets and experimental condition

Based on the observations about naked images collected from Internet, we categorized naked imagery into three types as illustrated in Fig.3: Type 1 includes naked images of whole naked body; Type 2 includes naked images cropped by middle size of

naked; Type 3 includes naked images zoomed into a part of naked body.

(a) (b) (c)

Fig. 3. Three types of real-world naked images. (a) Type 1. (b) Type 2. (c) Type 3.

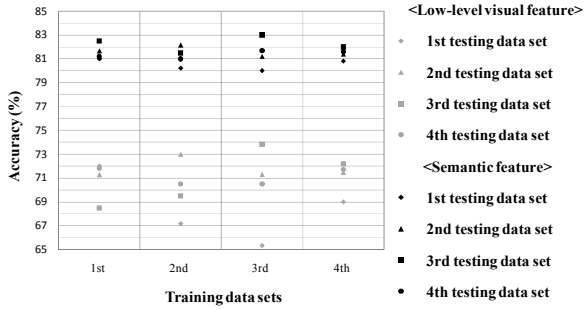

Fig. 4. The accuracy graphs of testing data sets with respective to the 4 training data sets with low-level visual features [3] and proposed semantic features.

The training and the testing samples were collected from the Internet. There were 4 training data sets and 4 testing data sets: 1st training and testing data sets consisted of images including Type 1; 2nd training and testing data sets consisted of images including Type 2; 3rd training and testing data sets consisted of images including Type 3; 4th training and testing data sets consisted of images including Type 1, 2, and 3 at the same ratio. In each data set, there were 300 images that belonged to 'naked' and 300 images that belonged to 'bikinied'.

To evaluate the effect of semantic features, we performed classification experiment with the testing data sets with respective to each training data set. For low-level visual features to be compared with the semantic features, Color Layout (CL), Color Structure (CS), Scalable Color (SC), Homogeneous Texture (HT), and Edge Histogram (EH) of MPEG-7 descriptors were used. For the semantic features, seven semantic concepts ('body', 'breast', 'genital', 'bottom', 'bikini', 'brassiere', 'panty') devised in section 2,1 were used. Binary SVM was used as a classifier in both low-level visual features and the semantic features.

4.2. Experimental results

To verify the usefulness of the adult semantic features compared to low-level visual features, we measured the accuracy of classifying naked images from bikinied images. The accuracy results for using the proposed 7 semantic features ranged from 80% to 83% as shown in Fig. 4. Then, the accuracy results for using low-level visual features ranged from 65% to 74%. According to these results, we could know two facts. Firstly, the proposed semantic features are more useful to classify the naked images from bikinied images than low-level visual features. Then, the proposed semantic features are also robust to classify the testing images which are unseen in training data set. Because when semantic features are used to classify unseen testing data sets, the fluctuation of classification performance for unseen testing data sets is lower than using low-level visual features.

In order to explain why the semantic features had robustness for the unseen testing images, we measured the proposed ADR. In the table 1, the low-level visual feature had low ADR values while the proposed semantic features had high ADR. Then, the SVMs trained with 4th training data sets have simultaneously higher ADR values, because 4th training data sets include all 3 type images. Namely, if training data sets consisted of all type images, then the classification performance can be stable for the testing images. However, in real world, the training images couldn't be composed with all type images. So, in real world, the semantic features could have more stable classification performance than low-level visual features.

Table 1. The ADR results according to used features

Training data sets	The naked image classification	
	Low-level visual features	Semantic features
1st	0.3	0.7
2nd	0.2	0.64
3rd	0.15	0.6
4th	0.47	1

5. CONCLUSIONS

In this paper, we proved that semantic features were useful to classify naked image and robust to the unseen testing images than low-level visual features. We proposed ADR measurement to evaluate the effect of used features related to the hyper-plane of SVM. Through experimental results, we showed the effect of semantic features with the proposed ADR. In further work, we additionally experiment to show the effect of semantic feature with the pose variation of naked images.

6. ACKNOWLEDGE

This work was supported by the IT R&D program of MKE/KEIT (2009-F-054-01, Development of the Illegal and Objectionable Multimedia Contents Analysis and Filtering Technology).

7. REFERENCES

[1] **J. S. Lee**, Y.-M. Kuo, P.-C. Chung, and E.-L. Chen, "Naked image detection based on adaptive and extensible skin color model," *Pattern Recognition*, vol. 40, pp. 2261-2270, 2007.

[2] J.-L. Shih, C.-H. Lee, and C.-S. Yang, "An adult image identification system employing image retrieval technique," *Pattern Recognition Letters*, vol. 28, pp. 2367-2374, 2007.

[3] W.I. Kim, H.-K. Lee, S. J. Yoo, and S. W. Baik, "Neural network based adult image classification," *ICANN 2005*, pp. 481-486, 2005.

[4] S.M. Kim, H.S. Min, J.H. Jeon, Y. M. Ro, and S.W. Han, "Malicious content filtering based on semantic features," *The ACM International Conference Proceeding 2009*, Nov. 24-26, 2009.

[5] S.J. Yang, S.-K. Kim, and Y. M. Ro, "Semantic home photo categorization," *IEEE Tran. On Circuits and Systems for Video Technology*, vol. 17, no. 3, Mar., 2007.

[6] M. Boutell, A. Choudhury, J. Luo, C.M. Brown, "Using Semantic Features for Scene Classification: how Good do they Need to Be?" *IEEE International Conference on Multimedia and Expo (ICME 2006)*, Jul. 9-12, 2006.

[7] M. Naphade, J. R. Smith, J.Tesic, S. F. Chang, W. Hsu, L. Kennedy, A. Hauptmann, and J. Curtis, "Large-scale concept ontology for multimedia," *IEEE Multimedia*, vol. 13, no. 3, pp. 86-91, Jul.-Sep., 2006.

[8] H. T. Lin, C. J. Lin, and R. C. Weng, "A note on Platt's probabilistic outputs for support vector machines," Dept. Comp. Sci., National Taiwan Univ., 2003 ([online]. Available: http://www.csie.ntu.deu/tw/~cjlin/papers/plattprob.ps, Tech. Rep).

[9] A. R. Webb, *Statistical Pattern Recognition*, second ed. John Wiley & Sons, 2002.