Recognition of Pornographic Web Pages by Classifying Texts and Images

Weiming Hu, Ou Wu, Zhouyao Chen, Zhouyu Fu, and Steve Maybank, *Senior Member*, *IEEE*

Abstract—With the rapid development of the World Wide Web, people benefit more and more from the sharing of information. However, Web pages with obscene, harmful, or illegal content can be easily accessed. It is important to recognize such unsuitable, offensive, or pornographic Web pages. In this paper, a novel framework for recognizing pornographic Web pages is described. A C4.5 decision tree is used to divide Web pages, according to content representations, into continuous text pages, discrete text pages, and image pages. These three categories of Web pages are handled, respectively, by a continuous text classifier, a discrete text classifier, and an algorithm that fuses the results from the image classifier and the discrete text classifier. In the continuous text classifier, statistical and semantic features are used to recognize pornographic texts. In the discrete text classifier, the naive Bayes rule is used to calculate the probability that a discrete text is pornographic. In the image classifier, the object's contour-based features are extracted to recognize pornographic images. In the text and image fusion algorithm, the Bayes theory is used to combine the recognition results from images and texts. Experimental results demonstrate that the continuous text classifier outperforms the traditional keyword-statistics-based classifier, the contour-based image classifier outperforms the traditional skin-region-based image classifier, the results obtained by our fusion algorithm outperform those by either of the individual classifiers, and our framework can be adapted to different categories of Web pages.

Index Terms—Web pages, pornographic texts, pornographic images, data fusion, recognition.

1 INTRODUCTION

The World Wide Web, as a global data center, allows people all over the world to share and exchange their information. Although the World Wide Web brings convenience, Web pages with harmful or illegal content are widely available. Among these, pornographic Web pages are the most common. Recognition of pornographic pages is very important for the proper development of Web resources and culture. However, recognizing pornographic Web pages is an extremely hard problem. One of the biggest problems is that different people have different definitions of pornography. The American laws [45], [46] define pornographic content as content that, taken as a whole, appeals to a prurient interest in nudity, sex, and so forth in the opinion of the average person, applying contemporary community standards. In addition, the type of the Web content itself changes over time and, accordingly, the techniques for recognizing pornographic Web pages must be adapted or improved over time.

Recognition of pornographic Web pages has attracted much attention. Besides laws [45], [46], platforms for content selection [2], [30], content ratings [3], action plans [4], and so forth, in recent years, some pioneering attempts have been made to recognize such Web pages automatically.

- *W. Hu, O. Wu, Z. Chen, and Z. Fu are with the National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences, PO Box 2728, Beijing, 100080, P.R. China. E-mail: {wmhu, owu, zychen82, zyfu}@nlpr.ia.ac.cn.*
- *S. Maybank is with the School of Computer Science and Information Systems, Birkbeck College, Malet Street Bloomsbury, London WC1E 7HX, UK. E-mail: sjmaybank@dcs.bbk.ac.uk.*

Manuscript received 21 Aug. 2006; revised 11 Dec. 2006; accepted 19 Dec. 2006; published online 20 Mar. 2007.
Recommended for acceptance by J. Luo.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0615-0806. Digital Object Identifier no. 10.1109/TPAMI.2007.1133.

1.1 Related Work

Methods for recognizing pornographic Web pages can be classified into text-based recognition, image-based recognition, and text and image combination-based recognition.

1.1.1 *Recognition of Pornographic Texts*

Current approaches for recognizing pornographic texts are mainly based on keyword matching and statistics and on text classification. Lee et al. [5], [6] count the frequencies with which keywords appear in a text. The frequencies, together with the relevant Web page features, are used as the input to the Kohonen self-organizing neural network (KSOM). After the learning stage is completed, the KSOM is used to determine whether a text can be classified as pornographic. Du et al. [7] extract feature vectors from pornographic and normal texts and save them in a database. To test a new text, its feature vector is extracted and matched to each of the saved feature vectors. Ho and Watters [8] use the Bayes classifier to recognize pornographic texts. They not only consider the influence of different words on the weights of the Bayes network but also assign different weights to the same words when they appear in different Web page components such as title, meta, and body.

In our opinion, there are three major problems in the recognition of Web pornographic texts:

- "**Overblocking**." Normal texts related to pornographic topics such as sex-related health and culture may be mistakenly classified as pornographic.
- "**Misspelled**." In antirecognition techniques, keywords in pornographic texts are misspelled deliberately in order to prevent recognition.
- "**Wordlist**." At present, there is no sufficient and practical list of keywords for recognizing pornographic texts.

1.1.2 Recognition of Pornographic Images

Automatic recognition of pornographic images has been studied by some researchers. Current methods can be classified as model-based, feature-based, and region-based.

Model-based recognition. Model-based methods emphasize the use of a human body model. Fleck and Forsyth [9], [10], [11], [12], [13], [14] segment the skin color pixels in an image using color and texture information and then find all connected columned skin regions that are candidates for trunks and limbs. These skin columns are combined subject to constraints derived from a geometrical model of the human body. If the combination can form the shape of a human body, the image is treated as pornographic. Ioffe and Forsyth [15], [16], [17] use the Monkatrol sampling method to learn the geometric relations between the different parts of the human body, extending the work in [9], [10], [11], [12], [13], [14].

The disadvantage of the model-based methods is that the human body has a complicated structure. It is difficult to consider all the possible relative positions of the parts of the body. Furthermore, these methods have a high-computational complexity, making them unsuitable for widespread use.

Feature-based recognition. Feature-based methods emphasize the extraction of features in pornographic images. Wang et al. [18], [19] use wavelet analysis to extract features from an image. The Daubechies wavelet transformation, the normalized central moments, and the color histogram are used to form semantic-matching vectors to classify images. There is no special emphasis on detecting skin. Skin detection is included in the analysis of the color histogram.

Region-based recognition. Region-based methods extract features for recognition based on the detected skin regions. Jiao et al. [20] use the proportion of skin area in the image and the area of the largest connected skin region to form the feature vectors. These feature vectors are used to train a support vector machine (SVM) with the radial basis function (RBF) Kernel. Liang et al. [21] extract the total skin area, the number of colors in skin regions, the number of connected skin regions, and the number of columned regions among the connected skin regions, together with the length, width, and moments of the largest connected skin region, to form a feature vector. The papers [22], [23], [24] compare the contributions of different features for recognizing pornographic images. It is concluded that the features that make the most important contributions are the total area of the skin regions, the area of the largest connected skin region, the number of skin regions, and the number of colors in skin regions. Arentz and Olstad [25] extract a series of features from each connected skin region: color, texture, and shape, together with the normalized center and the area of the region. A skin region, rather than the whole image, is used as a unit for recognition. Zheng et al. [26] construct a Maximum Entropy Model for the distribution of skin color to segment the skin regions in an image. They extract features from the skin regions and use a multilayer perception for training and recognition.

1.1.3 Text and Image Combination-Based Recognition

Some researchers begin to combine textual information and image information to recognize pornographic Web pages. Hammami et al. [27], [28] develop a text classifier and an image classifier and then connect the two classifiers in series to recognize pornographic Web pages. However, the image classifier is not very accurate because it uses only the ratio of skin pixels to all pixels in the Web page as the recognition feature. Jones and Rehg [29] combine a skin-model-based image detector and a text detector to detect pornographic Web pages and find that this combination improves the detection rate. The limitation is that the combination is based on the OR operation of the two detectors, that is, if either of the two detectors find pornographic content, then the Web page is classified as pornographic. This combination is not based on a proper calculation of probabilities. The result is an increased false acceptance rate.

1.2 Our Work

In this paper, we propose a novel framework for recognizing pornographic Web pages. In the framework, texts, images, and content representations of pages are jointly considered. Our framework is original in the following ways:

- We divide Web pages, according to content representations, into three categories: continuous text pages, discrete text pages, and image pages (the detailed definitions of the three categories of pages are given in Section 2). We propose a C4.5 decision tree algorithm for assigning Web pages to these categories.
- We propose an algorithm for recognizing pornographic continuous texts. Instead of a traditional keyword-statistics-based algorithm [5], [6], [7], [8], a CNN (Cellular Neural Network)-like word net is constructed to represent the semantic connections between keywords. Better results are obtained.
- We propose a naive Bayes classifier for recognizing pornographic discrete texts. The classifier outputs the probability that an input discrete text is pornographic. These probabilities are directly used in the fusion of Web information.
- We propose a new approach for estimating the parameters of the mixture model of skin colors. A multidimensional histogram is used to group the color vectors of neighboring pixels and, then, the Gaussian mixture model of the histogram is built. Our approach speeds up Gaussian mixture skin modeling without compromising the accuracy of the resulting model.
- In the contour-based algorithm for recognizing pornographic images, the image plane can be partitioned into any number of rectangular blocks. Experimental results show that the contour-based algorithm outperforms the traditional skin-region-based algorithm.
- We propose a fusion algorithm, based on Bayes theory, to fuse the recognition results from texts and images. Experimental results show that the fusion algorithm outperforms either of the individual classifiers.

The remainder of this paper is organized as follows: Section 2 gives an overview of our framework. Section 3 introduces the C4.5 decision tree algorithm for dividing Web pages. Sections 4, 5, and 6 present, respectively, the continuous text classifier, the discrete text classifier, and the image classifier. Section 7 describes the algorithm for fusing the recognition results for texts and images. Section 8 demonstrates experimental results. Section 9 summarizes the paper.

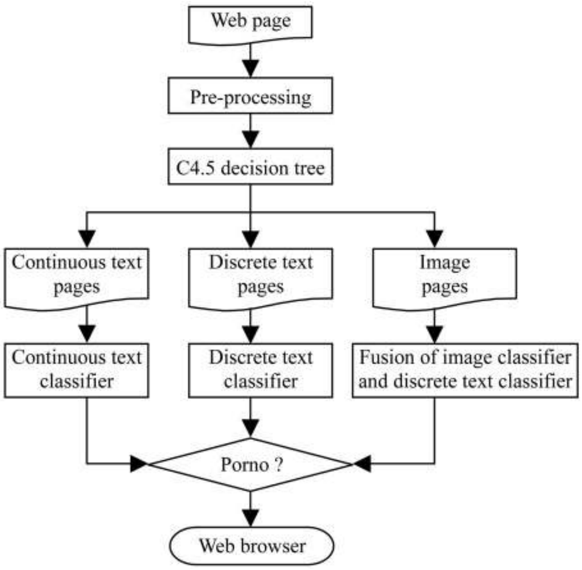


Fig. 1. Overview of our framework.

2 Overview of Our Framework

Existing approaches serialize text-based recognition and image-based recognition [27], [28], [29]. In this way, the error rates and time complexities are summed up from one stage to the next. In this paper, we divide Web pages into some categories and the recognition methods are adapted to each category. In this way, the complexities and the error rates from the classifiers do not accumulate.

Kleinberg [37] divides Web pages into Hub pages and Authority pages for evaluating page weights [36], where Hub pages supply the set of hyperlinks pointing to Authority pages. We have observed that Web pages can be divided, according to content representation, into the following categories:

- **Continuous text pages**. "Continuous texts" are defined as those in which there are full semantic and logical associations, for example, as found in articles. Continuous text pages mainly contain continuous texts, where very few images that are valid to represent the Web contents are involved.
- **Discrete text pages**. "Discrete texts" are defined as those that provide guidance or explanation. Discrete texts have less semantic associations than continuous texts. Discrete text pages mainly contain discrete texts, acting as navigators or head pages.
- **Image pages**. Image pages mainly contain image information. Their contents can be determined only by the image information. There are usually some discrete texts to explain the images in image pages.

In our framework, Web pages are divided into the above three categories, each of which is handled by the corresponding classifier. Fig. 1 shows an overview of our framework. The preprocessing rapidly extracts, according to content representations, texts and images from the Web page, discarding those inconsistent with the theme of the page, for example, banners and advertisements. The C4.5 decision tree algorithm classifies Web pages into the three categories. The continuous text classifier is used to detect pornographic content in continuous text pages. The discrete text classifier is used to detect pornographic content in discrete text pages. The image classifier is used to detect pornographic content in image pages. As discrete texts usually exist in image pages to provide captions or explanations for the images, the discrete text classifier is applied to classify the discrete texts among these pages. The recognition result is determined by the fusion of the results obtained by the image classifier and the discrete text classifier. The recognition result is transmitted to the Web browser for further handling.

3 Division of Web Pages

In our algorithm for dividing Web pages into the three categories, features of Web content representations are first extracted and Web pages are then divided using the extracted features.

Feature extraction. The time taken to assign a Web page to one of the three categories must be acceptably small. The following easily extracted page features are used:

- whether the URL contains terms like "main" or "index" indicating that the Web page is an indexing one,
- the number of characters of plain texts (not including hyperlink texts) in the Web page,
- the number of characters of hyperlink texts in the Web page,
- the number of images that contain more than 50,000 pixels,
- the number of images that contain 10,000 to 50,000 pixels, and
- the number of images that contain less than 10,000 pixels.

Classification. The C4.5 decision tree algorithm [35] is used to divide Web pages into the three categories, as the algorithm is accurate for classification and the model it uses is easily understood. The decision tree is constructed using a set of sample pages, and the decision tree is then applied to the test pages. During the training process, the set of sample pages is repeatedly partitioned into subsets until the sample pages in each subset all belong to the same category. In each step of the partitioning, the feature that produces the maximum information gain, as measured by the entropy, is selected as the reference for the partitioning. The sample pages that have the same value of the selected feature are assigned to the same subset.

Let $\Re$ be the number of categories of pages (in this paper, $\Re$ equals 3). For a set D of sample pages, let $p(D,r)$ be the ratio of the number of pages belonging to category r to the number ($|D|$) of pages in D. Then, the entropy of the $p(D,r)$, $1 \leq r \leq \Re$, is defined as

$$E(D) = -\sum\nolimits_{r=1}^{\Re} p(D,r) \times \log_2 p(D,r). \qquad (1)$$

Let T be a feature that can take any one of ν distinct values. Let D_i be the subset of sample pages corresponding to value i of T when feature T is selected as a reference for partitioning D. The information gain corresponding to the partitioning of D according to the feature T is defined as

$$G(D,T) = E(D) - \sum\nolimits_{i=1}^{\nu} \frac{|D_i|}{|D|} \times E(D_i), \tag{2}$$

where $|D_i|$ is the number of pages in the subset D_i, and $E(D_i)$ is defined in the same way as $E(D)$. The function $G(D,T)$ is very sensitive to the value of ν, so the ratio of information gain is used instead:

$$\lambda(D,T) = \frac{G(D,T)}{-\sum_{i=1}^{\nu} \frac{|D_i|}{|D|} \times \log_2\left(\frac{|D_i|}{|D|}\right)}. \tag{3}$$

The feature Γ satisfying

$$\Gamma = \arg\max_{T}(\lambda(D,T)) \tag{4}$$

is selected as the reference for this step of the partitioning.

4 Classification of Continuous Texts

Existing approaches for recognizing pornographic texts compile a list of keywords and then count the frequencies with which the keywords appear in a text to form a feature vector that summarizes the statistical features of the text. Some of these keywords appear in many legitimate texts. Statistical features alone are not enough to correctly classify these texts. In addition, these approaches fail if the keywords are deliberately misspelled. Semantic information is needed to handle these problems.

4.1 Semantic Feature Analysis

The same words may have different meanings in pornographic texts and in normal texts. Contextual information can be used to determine whether such words indicate the presence of pornographic content. Some words do not contain any pornographic meaning by themselves, but they do indicate the presence of pornographic content when they are combined with other words. Based on these considerations, we divide keywords into the following classes:

- **Obvious keywords**. These keywords mainly appear in pornographic texts rather than in normal texts and have pornographic meanings by themselves.
- **Hidden keywords**. These keywords do not have pornographic meanings by themselves, but they appear in pornographic texts with high probability.
- **Logical keywords**. These keywords are divided into two subclasses. One subclass consists of keywords that provide legitimate information in normal texts and pornographic information in pornographic texts. For example, "breast" is a keyword of this type. The second subclass consists of keywords that contain pornographic information only if they are associated with certain other specific words.

Many existing approaches only use obvious keywords and the first subclass of logical keywords. In fact, hidden keywords and the second subclass of logical keywords can greatly help the recognition of pornographic texts.

We consider the fact that there are associations between the words and these associations represent semantic relations between these words. For example, when we read a word in an article, we may associate it with other semantic-related words. This mechanism enlightens that the three classes of keywords and their semantic relations can reasonably represent the semantic features of the pornographic texts. In the Section 4.2, we construct a CNN-like word net to describe semantic features of pornographic texts.

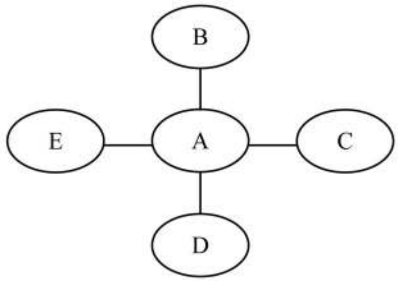


Fig. 2. A word cell and its adjacent cells.

4.2 CNN-Like Word Net

A CNN [31], [32] is formed by connecting a number of cells, where each cell is a processor with multiple inputs and a single output. Each cell is connected to a few adjacent cells. The information processing in a CNN is implemented by the local information exchange between adjacent cells. Each cell is characterized by an internal state variable. The output of each cell at the next state is determined by its current internal state, its current output, and the outputs of its adjacent cells.

To construct a CNN-like word net, we define a cell for each keyword. The associations between keywords are represented by connections between the corresponding cells. Fig. 2 shows a cell and its adjacent cells, where cell "A" is connected with cells "B," "C," "D," and "E." When a keyword is read, the cell corresponding to this keyword accepts it as an input. Then, the cell's state is changed according to its previous state and the states of its adjacent cells. The following parameters are used to describe each cell:

- the state of the cell,
- the positions where the corresponding keyword has appeared in the text,
- the number of times that the corresponding keyword has appeared in the text, and
- the number of times that the cell has been activated.

The state of a cell takes one of the three values: "sleep," "fallow," and "active." The state of "sleep" applies when the corresponding keyword has not appeared at the current stage. The state of "fallow" applies when the keyword has appeared but has not been activated. The state of "active" applies when the keyword has been activated. Only the information about the activated keywords is used as features for detecting pornographic texts.

When there is an input for a cell, the state of the cell is adjusted in the following ways:

- If the cell corresponds to an obvious keyword, no matter which previous state it is, its state is set to "active" and the number of times that it has been activated is increased by 1.
- If the previous state of the cell is "active," no matter which keyword class it corresponds to, its state remains to be "active" and the number of times that it has been activated is increased by 1.
- If the cell corresponds to a hidden or logical keyword and its previous state is "sleep" or "fallow," we check whether the parameters of its

adjacent cells satisfy a predefined rule—for example, most of its adjacent cells have been activated and one of the observed positions of each of the activated adjacent cells is close to one of the observed positions of the cell. If the rule is satisfied, then the state of the cell is set to "active" and its activated number is set to "1;" otherwise, the state of the cell is set to "fallow."

The rules for activating the cells of hidden and logical keywords can efficiently distinguish between legitimate sex-related texts and pornographic texts. In a legitimate sex-related text, although a hidden or logical keyword may appear, if the parameters of the cells adjacent to the cell corresponding to the keyword cannot satisfy the predefined rule, the cell corresponding to the keyword is not activated. This increases the probability that the text is not classified as pornographic.

The mechanism for activating cells can help solve the problem of misspelled keywords. Although a keyword is misspelled in a text (in fact, the keyword does not appear in the text), if the parameters of the cells adjacent to the cell corresponding to the keyword satisfy the predefined rule, the cell corresponding to the keyword can be activated according to the mechanism for activating cells. In other words, if there is a cell whose corresponding keyword does not appear in the text but its adjacent cells satisfy the predefined rule to activate the cell, it is checked whether there is a word in the text similar to the keyword. If so, it is assumed that the keyword is misspelled, and the corresponding cell is then activated. This process accords with the fact that the context of a misspelled word can be used to understand it correctly.

4.3 Algorithm

The algorithm for recognizing pornographic continuous texts includes construction of the CNN-like word net, feature extraction, training, and classification.

Construction of the CNN-like word net. Construction of the CNN-like word net includes the selection of a list of keywords and the definition of keyword associations. A list of keywords is constructed through the combination of extraction of statistical features of texts and manual selection. First, algorithms for selecting text features [40], such as the term-strength-based one, document-frequency-based one, mutual-information-based one, and information gain-based one, are used to select keyword candidates from sample pornographic texts. Then, the keywords are selected manually from the list of candidates and classified manually into "obvious," "hidden," or "logical." If two keywords appear close together in most sample pornographic texts, then there is a potential association between them. This association is confirmed manually.

Feature extraction. The major steps for extracting features from a text are summarized as follows:

Step 1. The states of all cells are initialized to "sleep."

Step 2. A word is read from the input text and matched to the keyword in each cell in the CNN-like net. If the word is a keyword, the corresponding cell's parameters recording the positions and the number of times that the keyword has been observed are updated and, then, Step 3 is carried out. If the word is not a keyword, Step 2 is repeated.

Step 3. If the cell has already been activated, then its number of activated times is increased by 1. Otherwise, it is checked whether the cell can be activated, depending on the parameters of the adjacent cells. If so, the cell's state is set to "active," and its number of activated times is set to "1;" otherwise, its state changes to or remains "fallow."

Step 4. If the state of the cell is changed, the states of its adjacent cells are adjusted according to the predefined rule. The whole net is adjusted iteratively until there are no further changes in the states of the cells.

Step 5. If all words in the text have been processed, go to Step 6; else, go to Step 2.

Step 6. The numbers of activated times for all active cells are combined to form a vector representing the semantic and statistical features of the input text.

Training and classification. We choose SVM as our classifier because of its good performance in text classification [33]. The SVM algorithm finds hyperplanes between different classes of the training data. The hyperplanes are used to classify the test data.

5 Discrete Text Classifier

The semantic associations in discrete texts can be ignored, so only the frequencies with which the keywords appear in discrete texts are used as the features for recognition. We use the Bayes classifier to classify discrete texts, as the Bayes classifier outputs probabilities that a text belongs to the categories rather than a certain result that the text belongs to a category, and the probabilities are used for the fusion of texts and images in later stages.

Let $C = \{c_1, c_2, \ldots, c_{|C|}\}$ be the set of text categories, where $|C|$ is the number of categories (In this paper, there are only two categories: the pornographic category c_1 and the normal category c_2.). Given an input vector $(a_1, \ldots, a_n)$ corresponding to a text, where n is the number of components in the vector, the classifier is used to determine the probability $P(c_j|a_1, \ldots, a_n)$ that the vector belongs to category c_j. An application of the Bayes rule yields the following equation:

$$P(c_j|a_1, \ldots, a_n) = \frac{P(a_1, \ldots, a_n|c_j)P(c_j)}{P(a_1, \ldots, a_n)}. \tag{5}$$

The term $P(c_j)$ is estimated by the frequency of the samples belonging to c_j in the training data. The estimation of $P(a_1, \ldots, a_n|c_j)$ is difficult because the data are sparsely distributed in a high-dimensional space. Considering that keywords in discrete texts are relatively independent, we use the naive Bayes classifier [34] to handle the problem.

The naive Bayes classifier assumes that the feature vector components are independent of each other, so it is represented by

$$P(c_j|a_1, \ldots, a_n) = \frac{P(c_j)\prod_{i=1}^{n} P(a_i|c_j)}{P(a_1, \ldots, a_n)}. \tag{6}$$

The estimation of $P(a_i|c_j)$ needs much less training data than the estimation of the joint probability $P(a_1, \ldots, a_n|c_j)$.

Let $W = \{w_1, w_2, \ldots, w_{|W|}\}$ be the set of keywords, where $|W|$ is the number of keywords, and let $T_j = \{t_{j,1}, t_{j,2}, \ldots, t_{j,|T_j|}\}$ be the set of training texts for category c_j, where $|T_j|$ is the number of texts in set T_j. The naive Bayes algorithm for classifying texts is given as follows:

Step 1. The probability $P(w|c_j)$ of keyword w given category c_j is calculated by

$$P(w|c_j) = \frac{1 + \sum_{i=1}^{|T_j|} N(w, t_{j,i})}{|W| + \sum_{s=1}^{|W|} \sum_{i=1}^{|T_j|} N(w_s, t_{j,i})}, \tag{7}$$

where $N(w, t_{j,i})$ is the number of occurrences of keyword w in text $t_{j,i}$.

Step 2. For a new text t, the probability $P(c_j|t)$ of c_j conditional on t is calculated by

$$P(c_j|t) = \frac{P(c_j) \prod_{k=1}^{|W|} P(w_k|c_j)^{N(w_k,t)}}{\sum_{r=1}^{|C|} P(c_r) \prod_{k=1}^{|W|} P(w_k|c_r)^{N(w_k,t)}}, \tag{8}$$

where

$$P(c_j) = |T_j| / \sum_{l=1}^{|C|} |T_l|. \tag{9}$$

Step 3. It is checked whether the probability $P(c_1|t)$ that the input text t belongs to the pornographic category c_1 exceeds a predefined threshold. If so, the text t is classified as pornographic; otherwise, the text is classified as normal.

We explain three points here: 1) The above threshold can be estimated statistically by finding the threshold that corresponds to the best recognition result for testing a large set of sample discrete texts. 2) As the probability that a legitimate sex-related discrete text belongs to the pornographic category is generally lower than the probability that a pornographic discrete text belongs to the pornographic category, a suitable choice of the threshold ensures that legitimate sex-related discrete texts can be distinguished from pornographic discrete texts. 3) A discrete text page is always a navigator or a head page. The results of classifying the Web pages hyperlinked with the discrete text page can be used to further determine whether the discrete text page is a legitimate sex-related one or pornographic.

Besides the detection of pornographic content in discrete texts, the probabilities output by the classifier can be used for the fusion of texts and images, as described in Section 7.

6 Recognition of Pornographic Images

In the image Web pages, the text information, if any, is not sufficient to detect pornographic content. It is necessary to analyze the images. The first step in detecting pornographic images is skin detection. In this section, we introduce first a histogram-based mixture model for skin detection and, second, a contour-based algorithm for recognizing pornographic images.

6.1 Skin Detection

There is a plethora of skin detection methods. One popular category of the methods is to use parametric models such as Gaussian and Gaussian mixture models for skin colors [29], [38], [39]. The expectation-maximization (EM) algorithm, which is used to find mixture model parameters, has a high computational cost. This hinders the application of mixture models to large databases of skin pixels. In this paper, a multidimensional histogram is used to represent the distribution of skin pixels. The parameters of the skin mixture model are estimated using the histogram bins in order to reduce the run time.

Histogram representation of skin samples. Note that many data points are redundant in the skin samples. As the data points with similar color values have similar color probabilities, we use a multidimensional histogram to represent the data set and treat the data points falling on the same bin as a single data point with multiple occurrences.

Let the histogram be $H = \{\{h_1, n_1\}, \{h_2, n_2\}, \ldots, \{h_M, n_M\}\}$, where M is the total number of histogram bins, h_i is the center point of the ith bin, and n_i is the number of data points in the ith bin.

Histogram-based mixture skin modeling. In color spaces, skin color has a distribution with multiple peaks. Compared with a single Gaussian model, a mixture model with multiple Gaussians better describes the skin color distribution.

A mixture model Θ, which is the weighted sum of individual density functions, is used to describe the constructed skin histogram. Let α_j be the weight of the jth density function $p_j(h|\theta_j)$, where θ_j is a parameter vector specifying the jth density function, and let K be the number of density functions. The mixture mode Θ is then

$$p(h|\Theta) = \sum_{j=1}^{K} \alpha_j p_j(h|\theta_j). \tag{10}$$

We add a set of random variables $Y = \{y_1, y_2, \ldots, y_M\}$, each of which takes a value in the range $1, 2, \ldots, K$. The probability $p(y_i = j)$ describes the probability that h_i is generated from the jth Gaussian. The model parameters are estimated by maximizing the expectation of the log-likelihood function:

$$F(y_1, y_2, \ldots, y_M) = \sum_{i=1}^{M} \log(p(h_i, y_i|\Theta)) \tag{11}$$

using the EM algorithm. The expectation of the log-likelihood function is

$$E_y[F] = \sum_{j=1}^{K} \sum_{i=1}^{M} n_i \log(\alpha_j p_j(h_i|\theta_j) p(y_i = j|h_i, \Theta)). \tag{12}$$

In the following, $p(y_i = j|h_i, \Theta)$ is rewritten as $p(j|h_i, \Theta)$ for simplicity. The parameters are initialized randomly and iteratively updated by the following E-step and M-step.

The E-step identifies the expectation by computing the posterior probability $p(j|h_i, \Theta)$ for each Gaussian j, which is, following the Bayes rule, given by

$$p\left(j|h_i, \Theta^{(t-1)}\right) = \frac{\alpha_j^{(t-1)} p_j\left(h_i|\theta_j^{(t-1)}\right)}{\sum_{k=1}^{K} \alpha_k^{(t-1)} p_k\left(h_i|\theta_k^{(t-1)}\right)}, \tag{13}$$

where t is the number of iterations.

The M-step updates the parameters by maximizing the expectation. The derivative of the right-hand side of (12) is taken with respect to α_j and set to 0. Then, the following equation is acquired:

$$\alpha_j^{(t)} = \frac{1}{N} \sum_{i=1}^{M} n_i p\left(j|h_i, \Theta^{(t-1)}\right). \tag{14}$$

The mixture weights are updated by this equation. Let the jth individual Gaussian be described by mean μ_j and covariance matrix Σ_j. Equations (13) and (14) are substituted into (12), and the derivatives of right-hand side of (12) are taken with respect to μ_j and Σ_j, respectively, and set to 0. Then, the following equations are acquired:

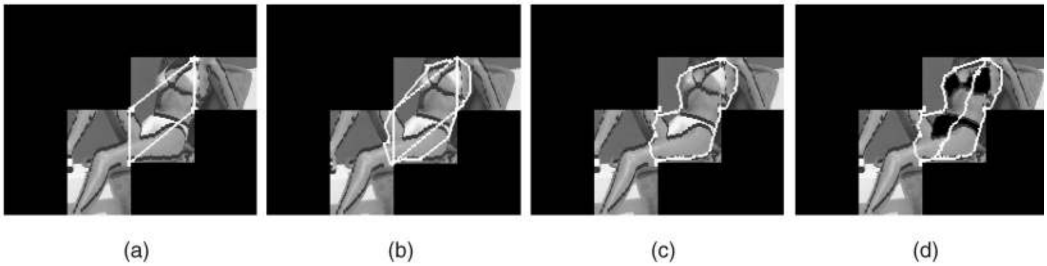


Fig. 3. Contour and local information [1]: (a), (b), and (c) Contour extraction process. (d) Local information.

$$\mu_j^{(t)} = \frac{\sum_{i=1}^{M} n_i h_i p(j|h_i, \Theta^{(t-1)})}{\sum_{i=1}^{M} n_i p(j|h_i, \Theta^{(t-1)})}, \tag{15}$$

$$\Sigma_j^{(t)} = \frac{\sum_{i=1}^{M} n_i \left(h_i - \mu_j^{(t)}\right)\left(h_i - \mu_j^{(t)}\right)^T p(j|h_i, \Theta^{(t-1)})}{\sum_{i=1}^{M} n_i p(j|h_i, \Theta^{(t-1)})}. \tag{16}$$

The parameters are updated iteratively by (13), (14), (15), and (16).

Histogram bin merging. One method for representing data points using a histogram is to subdivide the sample space at uniform scale by setting a fixed bin size for the multidimensional histogram. Considering the nonuniform nature of the skin data distribution, an alternative is to set bin size adaptively. Smaller bins are used where the data points are densely populated and larger bins where the data points have a sparse distribution.

The construction of the multidimensional histogram with adaptive bin size begins with the marginal histograms with a fixed bin size for each coordinate in the sample space. Adjacent bins are merged if the sum of data points is smaller than a predefined threshold. The resulting multidimensional histogram represents the joint distribution of marginal histograms after bin merging. This merging reduces the size of the data set, without compromising the accuracy of results, as illustrated in Section 8.3.1.

The above describes our color histogram-based Gaussian mixture model. It is noted that, before estimating the parameters of the mixture model, the fuzzy K-Means algorithm is used to cluster the skin pixels to give the initial values for the EM algorithm. This initialization speeds up the convergence of the EM algorithm. The fuzzy K-Means algorithm can better cluster data than the EM algorithm, but the EM algorithm is better for estimating the parameters of the Gaussian mixture model. The combination of the fuzzy K-Means algorithm and the EM algorithm ensures that the parameters of the Gaussian mixture model are evaluated accurately and rapidly.

In the following, we describe our approach for pornographic image recognition, which is based on the skin color pixel detection using the above described Gaussian mixture skin model.

6.2 Pornographic Image Recognition

The most popular approaches for recognizing pornographic images use feature extraction from segmented skin regions. These approaches include weak shape constraints associated with the human body. Consequently, they cannot classify accurately normal images rich in skin colors, such as face images and legitimate sex-related images of images of women wearing bikinis, swimsuits, and so forth. In this section, Yang et al.'s work [1] (our previous work) is extended to produce a generalized contour-based pornographic image recognition algorithm that uses effective human shape constraints to distinguish normal images rich in skin colors from pornographic images.

In the algorithm, the image plane is partitioned into "rectangular blocks" of equal size. The blocks containing many skin pixels are defined as Blocks of Interest (BOIs). The BOIs' corners, around which there are many skin pixels, are defined as Points of Interest (POIs). A "connected block region" is a four-connected set of BOIs. We find "the largest connected block region" (rather than the connected skin region) from which the contour of the object in the image is extracted.

Contour extraction. The contour extraction process is summarized in the following steps:

Step 1. The POIs on the border of the "largest connected block region" are connected along the border to form a closed curve A, as shown in Fig. 3a [1].

Step 2. Outside curve A, a curve parallel and close to curve A is drawn. The skin edge points are found within the region Φ between the parallel curve and curve A and fitted along curve A to form a closed curve B, as shown in Fig. 3b. (The approach for producing curve B is described as follows: In the left part of the region Φ, for the edge points with the same x-coordinate, only the left one is kept, that is, the others are removed and, then, the kept edge points are connected in the order of increase of the y-coordinate. The edge points in the right part of the region Φ are handled conversely.)

Step 3. Curve B is refined through adjusting the locations of the nonskin points in curve B, as shown in Fig. 3c. For a nonskin point in curve B, along the direction perpendicular to the nearest line in curve A, a skin edge point is found and used to update the nonskin point in curve B. The refined curve C represents the contour of the object in the image.

Step 4. Nonskin regions enclosed within curve C are detected, and some local information such as area and positions of the detected nonskin regions and their distributions in the two regions divided by the middle curve of the contour, as shown in Fig. 3d, is acquired as the features for determining whether the image is pornographic. (The approach for estimating the middle curve is described as follows: The two points on the contour with the maximum distance are found. The two points are linked to form a line segment. Each line perpendicular to the line segment intersects the contour to produce two points of intersection. We take the midpoint of the line segment linking the two points of intersection. These midpoints are linked orderly to form the middle curve. An example of such a curve is shown in Fig. 3d.)

We explain four points here:

1. As long as there exists a connected block region in the image, the contour can be extracted regardless of the pose of the object in the image. If there is no BOI in the image, no connected block region can be extracted, and the image is classified as normal.
2. The connected skin regions need not be extracted.
3. In legitimate sex-related images, although connected skin regions are broken by swimsuits, bikinis, and so forth, the connected block region can still be found and the contour of the object can be extracted. The features needed to distinguish legitimate sex-related images from pornographic images can then be extracted.
4. In our previous algorithm [1], the image plane is partitioned into 4×4 rectangular blocks. In this generalized algorithm, the image plane can be partitioned into any number of rectangular blocks.

Feature extraction. Similar to [1], the following image features, based on the extracted contour, are used to distinguish pornographic images from normal images:

- the numbers of skin pixels in all the "rectangular blocks" into which the image is partitioned, describing the distribution of skin pixels in the image,
- the three angles of the triangle formed by the centers of the three nonskin regions with the biggest areas in the two sides of the middle curve,
- the mean and covariance matrix of positions of all the detected nonskin pixels within the contour, describing the spread of the nonskin pixels,
- the ratio of the distances from the center of the largest nonskin region to the two endpoints of the middle curve, indicating the information of the position of the largest nonskin region within the contour, and
- the breadth to length ratio of the image.

The above features are combined to form a feature vector v, which is used for classifying images.

Image classification. Similar to [1], the nearest-center classifier is employed for classification. The feature vector of an input image is matched to the feature vectors of sample images. In the image classification method, sample images are grouped according to their numbers of BOIs, and feature vectors are matched only for images with the same number of BOIs. In this way, the number of matching is decreased, and the computation complexity is reduced.

The similarity between the feature vectors v_1 and v_2 of two images is defined to be

$$d(v_1, v_2) = 1 - \frac{v_1^T v_2}{\|v_1\| \times \|v_2\|}. \quad (17)$$

For the feature vector v of an input image that has n BOIs, from the known sample images with n BOIs, the sample image whose feature vector is most similar to vector v is found. The input image is assigned to the same class as the sample image.

7 Fusion of Images and Texts

In an image Web page, there may be many images, possibly with small amounts of discrete texts explaining the images. In this section, a method is described for fusing the information obtained from the images and the information obtained from the texts.

Note that a Web page has a semantic theme, and images and texts in the page express the same theme. Based on this observation, we assume the following a priori knowledge:

- Almost all the images that are large in size and central in position are pornographic in a pornographic Web page or normal in a normal page.
- The information in the discrete texts provides a prior classification of the images.

Then, the problem of classifying an image page becomes a problem of classifying a set of images, where the result of classifying the discrete text in the image page is used as a prior classification.

The output of our image classifier for an input image is whether this image is pornographic or not. We define two statistical features of the classifier:

- the probability (p_1) that a normal image is mistakenly classified as pornographic and
- the probability (p_2) that a pornographic image is mistakenly classified as normal.

The probabilities p_1 and p_2 can be estimated statistically by counting the number of images mistakenly classified by the image classifier in a large set of known images.

Suppose that there are N images in the test Web page and the result of classifying the N images using the image classifier is that N_1 images are classified as pornographic and N_2 ones as normal $(N_2 = N - N_1)$. Let $r = (N_1 \text{ pornographic images}, \ N_2 \text{ normal images})$. Let S represent the event that all the N images are pornographic and $\neg S$ the event that all the N images are normal. Then, the equations below are obtained:

$$p(r|S) = (1 - p_2)^{N_1} (p_2)^{N_2}, \quad (18)$$

$$p(r|\neg S) = (p_1)^{N_1} (1 - p_1)^{N_2}. \quad (19)$$

According to the Bayes rule, the following equations are obtained:

$$p(S|r) = \frac{p(r|S) \times p(S)}{p(r)}, \quad (20)$$

$$p(\neg S|r) = \frac{p(r|\neg S) \times p(\neg S)}{p(r)}. \quad (21)$$

We introduce a decision factor f, which is the ratio of the two posterior probabilities in (20) and (21):

$$f = \frac{p(S|r)}{p(\neg S|r)} = \frac{p(r|S) \times p(S)}{p(r|\neg S) \times p(\neg S)} = \frac{(1 - p_2)^{N_1} (p_2)^{N_2}}{(p_1)^{N_1} (1 - p_1)^{N_2}} \times \frac{p(S)}{p(\neg S)}. \quad (22)$$

If $f \geq 1$, the Web page is classified as pornographic.

The remaining problem is to evaluate the a priori probabilities $p(S)$ and $p(\neg S)$. The discrete texts around the images provide the information needed to obtain a prior classification of the images, that is, the a priori probability $p(S)$ is replaced with the probability P_t that the discrete text classifier classifies the text as pornographic. Accordingly, the probability $p(\neg S)$ is set equal to $1 - P_t$. Then,

$$f = \frac{(1-p_2)^{N_1}(p_2)^{N_2}}{(p_1)^{N_1}(1-p_1)^{N_2}} \times \frac{P_t}{1-P_t}. \tag{23}$$

If there is no text in the Web page, $p(S)$ is set to N_1/N and $p(\neg S)$ to N_2/N.

Two points should be mentioned here: 1) Our fusion between images and texts is based on semantic relations between images and texts, that is, images and texts on a page represent the same theme. Compared with traditional information fusion approaches [42], [43], [44] such as Bayes networks and Dempster-Shafer's theory, our fusion approach implements the semantic fusion of images and texts to some extent. 2) Sometimes, large images in a pornographic Web page are purposely broken into small pieces to avoid detection. In those cases, fusion of the recognition results of the small pieces of images can lead to a more accurate classification of the entire Web page.

The following examples show the rationality of our fusion algorithm:

1. **Example 1**. Suppose that $p_1 = p_2 = 0.2$ and $P_t = 0.5$. Then,

$$f = \frac{0.8^{N_1}0.2^{N_2}}{0.2^{N1}0.8^{N_2}} = \frac{(0.2\times 4)^{N_1}0.2^{N_2}}{(0.2\times 4)^{N_2}0.2^{N_1}} = \frac{4^{N_1}}{4^{N_2}}. \tag{24}$$

   If $N_1 > N_2$, then $f > 1$, that is, the Web page is classified as pornographic. If $N_1 < N_2$, then $f < 1$, that is, the Web page is classified as normal. This is quite reasonable in the intuitive view.
2. **Example 2**. Suppose that $p_1 = 0$, and $P_t = 0.5$. Then, the decision equation is transformed to

$$f = \frac{(1-p_2)^{N_1} p_2^{N_2}}{0^{N_1}}. \tag{25}$$

   If $N_1 > 0$, no matter how many images in the Web page are pornographic, the Web page is classified as pornographic. This means, if the probability that a normal image is mistakenly classified as pornographic by the image classifier is 0, as long as one image is pornographic, the Web page is classified as pornographic. This is consistent with real applications.

8 Experimental Results

All the above algorithms are implemented using Visual C++6.0 on the Windows XP platform. In the following, the performances of the continuous text classifier, the discrete text classifier, the image classifier, the algorithm for fusing Web page information, and the framework for classifying Web pages are evaluated in succession. All runtimes are measured on a Pentium-4 3.2-GHz computer.

8.1 The Continuous Text Classifier

To evaluate the performance of our continuous text classifier, 3,162 continuous texts were downloaded from the Internet, including 577 pornographic texts, 585 legitimate sex-related texts, and 2,000 normal texts not related to sex, as checked by hand. The normal texts cover 10 topics: "arts," "business," "sciences," "computer," "news," "shopping," "game & recreation," "society," "health," and "sports," and, for each topic, there are 200 corresponding Web texts. In contrast with other text databases [7] for recognizing pornographic content, we collected legitimate sex-related samples dealing with sexual health, sexual culture, sexual education, and so forth. We randomly selected 300 pornographic texts, 300 legitimate sex-related texts, and 1,000 normal texts not related to sex as training data. The remaining texts served as test data. There were 800 keyword candidates selected using the algorithms for extracting text features. Of these, 109 were selected manually as keywords, comprising 29 obvious keywords, 33 hidden keywords, and 47 logical keywords. There are 619 potential associations between keywords acquired automatically. Of these, 323 were manually selected as confirmed associations.

TABLE 1
Recognition Rates for the Three Schemes

	Pornographic texts	Legitimate sex-related texts	Normal texts not related to sex	Total
Scheme 1	93.1%	95.1%	100.0%	97.9%
Scheme 2	96.4%	97.5%	99.8%	98.8%
Scheme 3	97.8%	98.2%	100.0%	99.3%

The following different schemes for feature extraction are used to access the usefulness of the different sets of keywords:

- **Scheme 1**. Only obvious keywords and the first subclass of logical keywords are counted.
- **Scheme 2**. All the three classes of keywords are counted.
- **Scheme 3**. The keywords corresponding to the active cells in the CNN-like word net are counted.

Table 1 shows the results of the comparison between the classification rates of the three schemes for feature extraction tested in the same set of test samples. From the table, the following useful points are deduced:

- On comparing the results for Schemes 1 and 2, it is apparent that the introduction of hidden keywords and the second subclass of logical keywords increases the recognition rates noticeably for the pornographic texts and the legitimate sex-related texts due to the expansion of the list of keywords. However, as the keyword list contains hidden keywords and the second subclass of logical keywords, some normal texts not related to sex are classified, inevitably, as pornographic.
- The best recognition rates are obtained by using the CNN-like word net to extract features of texts. It can be declared that the CNN-like word net represents the semantic features of pornographic texts properly.
- The real pornographic texts have protean styles and contents. In order to reach higher classification rates, the keyword list should be enlarged.
- Only simple semantic relations between keywords are considered in the CNN-like word net-based approach, so some legitimate sex-related texts are still classified as pornographic.

We collected 150 misspelled pornographic texts to test the recognition rates of the above schemes. As shown in Table 2, for the misspelled pornographic texts, our CNN-like word net-based approach has a much higher recognition rate than

TABLE 2
Testing Misspelled Texts

	Misspelled texts
Scheme 1	56.0%
Scheme 2	58.0%
Scheme 3	87.3%

TABLE 3
Performance Comparison

	Correct classification rate	False acceptance rate
Our approach	97.8%	0.4%
Du's approach (t=0.18)	96.3%	0.5%
Du's approach (t=0.1)	97.1%	4.2%

Schemes 1 and 2. So, the semantic features of pornographic texts represented in the CNN-like word net are useful to solve the misspelled problem.

Du et al. [7] apply text classification to recognize pornographic texts. They tested their approach on a data set that contains normal texts, none of which are sex-related. Table 3 shows the comparisons, tested in our data set, between their approach and our CNN-like word net-based approach, where t in the table is a threshold in [7]. In the table, the correct classification rate is the fraction of the correct classified texts in the set of pornographic texts and the false acceptance rate is the fraction of normal texts incorrectly classified as pornographic. In the table, it is shown that the correct classification rate of our approach is higher than that of Du et al.'s and the false acceptance rate of our approach is lower than that of Du et al.'s. The main reason why our approach has better results is that it classifies legitimate sex-related texts more accurately.

To summarize, the experimental results with and without hidden keywords and the second subclass of logical keywords show that the introduction of hidden keywords and the second subclass of logical keywords is an efficient way to solve the wordlist problem. The experimental results in testing the legitimate sex-related texts show that semantic associations between keywords can solve the over-blocking problem efficiently. The experimental results in testing the misspelled texts show that our CNN-like word net-based approach can greatly improve the classification of misspelled pornographic texts.

8.2 The Discrete Text Classifier

A total of 165 keywords were selected for the discrete text classifier, and 3,000 discrete texts were collected from Web head pages and hypohead pages, including 1,000 pornographic texts and 2,000 normal texts covering many topics. The training data consisted of 500 of the pornographic texts and 1,000 of the normal texts. The remaining texts were used as test data.

Table 4 shows the performance of the naive Bayes classifier for classifying the test data under different choices of the threshold introduced in Section 5. In the table, it can be seen that the Bayes-based classifier accurately classifies the discrete texts. When the threshold is set to 0.8, the classifier reaches the best performance as a whole.

TABLE 4
Results of Testing Discrete Texts

Threshold (t)	Pornographic texts	Normal texts	Total
t=0.5	79.6%	94.9%	89.8%
t=0.6	78.8%	96.6%	90.7%
t=0.7	78.2%	97.5%	91.1%
t=0.8	77.8 %	98.5 %	91.6%
t=0.9	75.2%	99.2%	91.2%
t=0.95	73.0%	99.6%	90.7%

TABLE 5
Results of Continuous Text Classifier Testing Discrete Texts

Pornographic texts	Normal texts	Total
19.4 %	99.8%	73.0%

TABLE 6
Testing Legitimate Sex-Related Discrete Texts

Threshold (t)	Number of texts classified as pornographic	Number of texts classified as normal
t=0.5	10	30
t=0.6	8	32
t=0.7	4	36
t=0.8	2	38
t=0.9	2	38
t=0.95	0	40

Table 5 shows the performance of the CNN-like word net-based approach for classifying the discrete texts. In the table, it can be seen that the recognition rate of the continuous text classifier for the pornographic discrete texts is only 19.4 percent. This indicates that there are many differences between continuous texts and discrete texts. If the differences are not considered, in that the same classifier is used to classify these two categories of texts, then a low classification error cannot be obtained.

The average runtime of the Bayes classifier on the test data is 0.02 seconds while that of the CNN-like word net-based classifier is 0.09 seconds. As the associations between keywords are considered in the CNN-like word net-based classifier, its runtime is higher.

To test the influence of the threshold on the classification of legitimate sex-related discrete texts, 40 legitimate sex-related discrete texts were downloaded from the Internet. Table 6 shows the results of classifying these legitimate sex-related discrete texts with different choices of the threshold. In the table, it can be seen that, the higher the threshold, the more legitimate sex-related texts are classified as normal. When the threshold is set to 0.8, only two legitimate sex-related texts are mistakenly classified as pornographic, producing a satisfying result.

TABLE 7
Comparison of Gaussian Mixture Models

Number of Gaussians	Equal Error Rate (EER)			
	RGB(%)	HSV(%)	YCbCr(%)	r-g(%)
1	17.09	13.86	17.09	21.42
2	16.84	13.43	16.84	24.92
3	16.81	13.21	16.60	23.97
4	16.71	12.87	16.61	23.36
5	16.67	13.00	16.26	23.28
6	16.59	12.84	16.26	23.48
7	16.61	13.06	16.57	23.51
8	16.58	13.01	16.68	23.44
9	16.66	12.96	16.52	23.54
10	17.13	12.83	16.92	23.48
11	17.23	12.98	17.23	×
12	17.26	12.99	17.10	×
13	17.18	12.91	17.56	×
14	17.43	12.80	17.61	×
15	17.22	12.96	17.23	×

8.3 The Image Classifier

In this section, we evaluate the performance of the skin detection algorithm first and then the performance of the algorithm for recognizing pornographic images.

8.3.1 Skin Detection

Skin modeling. The choice of the number of Gaussians is important for mixture models. The number of Gaussians used in previous work on skin modeling varies greatly from 1 to 15 [38], [39], but, as yet, no comparisons have been made between mixture models with different numbers of Gaussians, partly due to the complexity of conventional mixture modeling.

In the experiments, we tested our histogram-based mixture models with different numbers of Gaussians using the most popular color spaces, namely, red, green, blue (RGB), Hue Saturation Value (HSV), YCbCr, and normalized r-g. The setup of the test is:

1. The test was carried out on the Compaq skin database [29], which contains about 5,000 images with skin color, including over 80,000,000 million labeled skin pixels.
2. The experiment included a fivefold cross validation. The database was divided into five disjoint subsets with equal size; in each round of the test, one subset was chosen as the test set and the remaining four subsets comprised the training set. The training and test process was repeated for five times and the results were averaged.
3. For each model, there are different false positive rates and different false negative rates associated with different thresholds used to determine whether pixels are classified as skin. For the best choice of the threshold, the false positive rate reaches equilibrium with the false negative rate. Therefore, the Equal Error Rate (EER) (where the false positive rate equals the false negative rate) is used as the performance index.

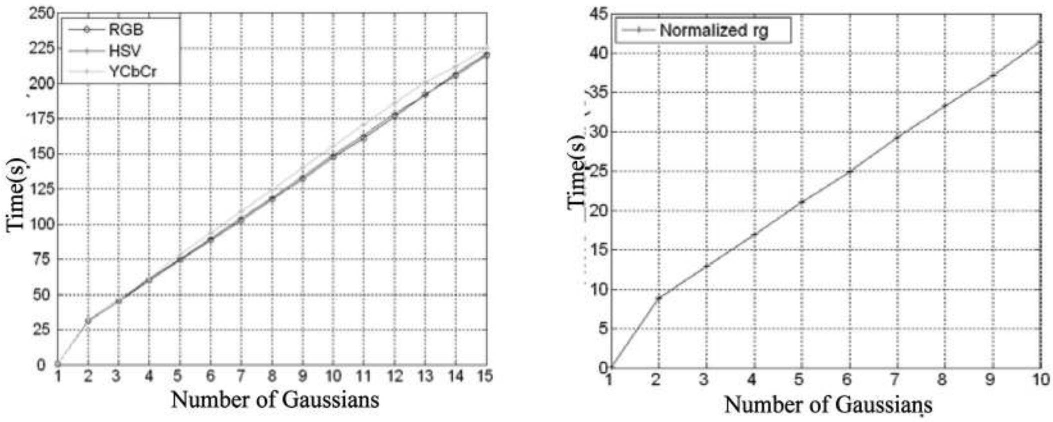
Table 7 compares the results with a varying number of Gaussians and for different choices of the color space. From Table 7, we have the following interesting discoveries:

1. For all the 3D color spaces, namely, RGB, HSV, and YCbCr, the Gaussian mixtures generally outperform the single Gaussian. The EERs decrease overall as the number of Gaussians increases from 1 to 5. The EERs change very slightly except for minor fluctuations when the number of Gaussians increases from 6 to 10. When the number of Gaussians is more than 10, the performance of the models reaches saturation and begins to downgrade due to overfitting.
2. For the 3D color spaces, the performance of YCbCr is comparable with that of RGB, as they are linearly related; the HSV space in which the chrominance and luminance information are separated provides the best accuracy in skin detection.
3. The single Gaussian model is not very weak. It is somewhat (instead of greatly) worse than the Gaussian mixtures except in the normalized r-g space.
4. The normalized r-g space, which was thought to be well suited for skin detection due to its robustness to changing illuminations, does not perform as well as expected. The reduction from a 3D color space to a 2D color space plays a role here, as useful information for skin classification is lost. For the normalized r-g space, there is little benefit in using more than one Gaussian.

Fig. 4 shows the average time spent for training skin models. In Fig. 4, it can be seen that the proposed method is efficient in training skin models. Given the same number of bins, comparable efficiencies are obtained for all the three 3D color spaces, as shown in Fig. 4a. The time spent on training the model with the maximum number of 15 Gaussians is less than 250 seconds. For the normalized r-g space, even less time is needed, as shown in Fig. 4b.

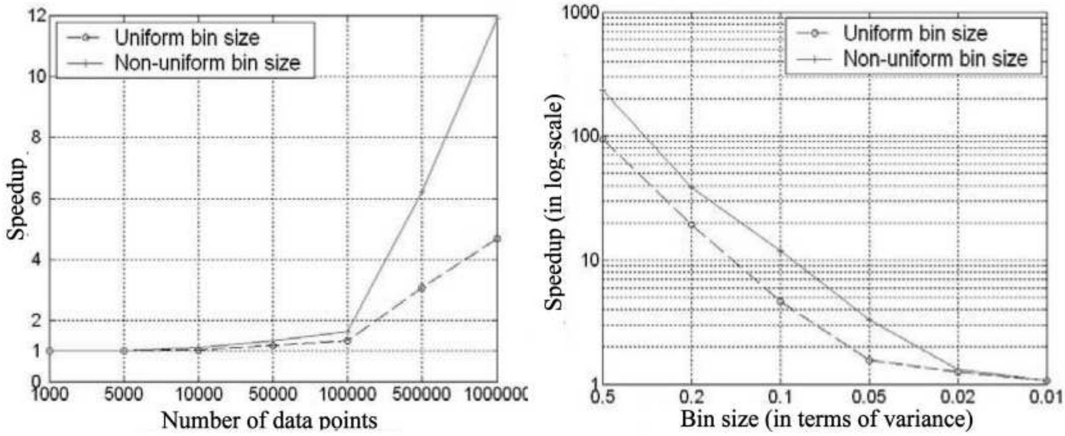
Test on synthetic data. Our method was further tested on synthetic data sets to show its efficiency. Random samples were generated from Gaussian mixtures in 3D space, with five Gaussians for each model by default. Fig. 5a shows the speedup ratios of our method compared with the conventional EM algorithm when the number of data points varies from 1,000 to 1,000,000. The dashed line shows the results of histogram bin modeling with a fixed bin size, whereas the solid line shows the results with adaptive bin size. Fig. 5b shows the speedup ratios with respect to different bin sizes (in terms of variance). In Fig. 5, we can see that the efficiency of our method improves considerably when the number of data points or the size of bins increases, and modeling with adaptive bin size brings a further speedup compared with fixed bin size.

Table 8 shows the relative errors in log-likelihood values between the estimated model and the ground truth. In Table 8, we can see that both strategies with fixed and adaptive bin sizes approximate the ground truth model well, given that the bin size is not too large. A bin edge size about 1/10 of variance seems to be a good trade-off between accuracy and efficiency.



(a) (b)

Fig. 4. Average time spent for building the skin models: (a) In RGB, HSV, and YCbCr spaces. (b) In normalized r-g space.



(a) (b)

Fig. 5. Speedup ratios of our method: (a) Effect of the number of data points. (b) Effect of bin sizes.

8.3.2 Pornographic-Image Recognition

The performance of the contour-based pornographic-image recognition algorithm is evaluated in our previous work [1], when the image plane is partitioned into 4×4 rectangular blocks (readers may read our previous paper [1] for more details). In the following, the performance of the contour-based algorithm is compared with the performance of a connected skin-region-based algorithm.

The connected skin-region-based pornographic-image recognition algorithms are the most popular in current research. We implemented a region-based algorithm based on the work in [21], [22], [23], [24]. Referring to the result of the comparison, in [21], [22], [23], [24], between contributions of the different features of the connected skin regions, the following features are selected for classifying images:

- ratio of the area of skin regions to the area of the image,
- average skin probability of all pixels in the image,
- number of connected skin regions,
- area ratio of the largest skin region to all skin regions,
- image width, and
- image height.

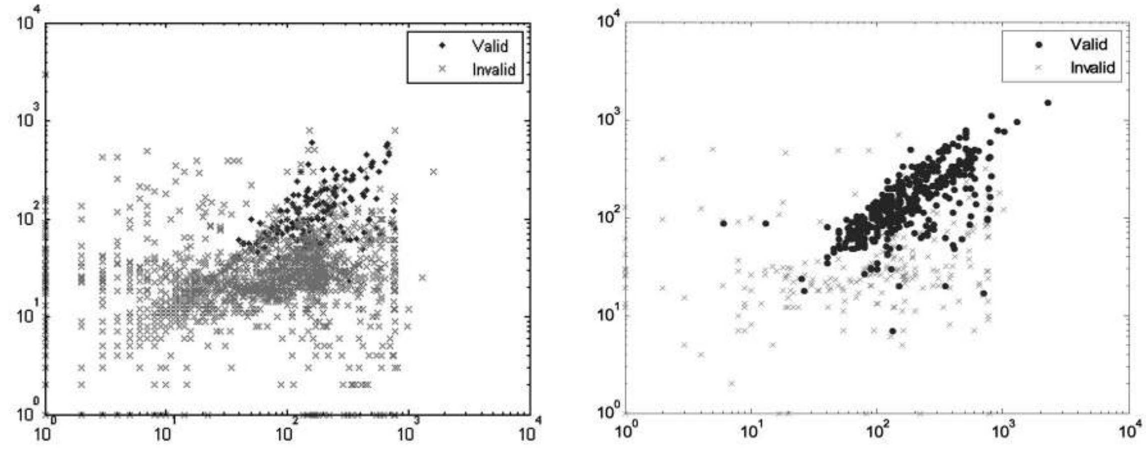
In real situations, images of women wearing bikinis or swimsuits or close range images of faces are often mistakenly classified as pornographic. Three sets of images are constructed: a set of pornographic images, a set of images of women wearing bikinis, swimsuits, and so forth (legitimate sex-related images), and a set of face images, where each set contains 2,000 images. Half of the images in each set are used as training samples and the remaining images as test samples. Table 9 shows the comparison between the correct classification and false acceptance rates of the contour-based algorithm and the correct classification and false acceptance rates of the region-based algorithm, tested in the three sets of images. In Table 9, it can be seen that, compared with the contour-based algorithm, for the pornographic images and the legitimate sex-related images (the images of women wearing bikinis, swimsuits, and so forth), the correct classification rates of the region-based algorithm are only slightly decreased, but the false

TABLE 8
Relative Errors in Log-Likelihood Values

Bin size	0.5σ	0.2σ	0.1σ	0.05σ	0.01σ
Fixed bin size	0.06	0.05	0.04	0.04	0.04
Adaptive bin size	0.12	0.06	0.05	0.04	0.04

TABLE 9
Comparison between the Contour-Based Algorithm and the Region-Based Algorithm

	Pornographic images		Bikini images		Face images	
	Contour-based	Region-based	Contour-based	Region-based	Contour-based	Region-based
Correct classification rate	92.8%	89.3%	94.3%	91.4%	93.7%	86.1%
False acceptance rate	6.0%	10.1%	2.8%	12.5%	6.5%	13.6%



(a) (b)

Fig. 6. Distributions of sizes of valid and invalid images: (a) GIF images. (b) JPEG images.

acceptance rates of the region-based algorithm are much increased. For the face images, the correct classification rate of the region-based algorithm is much decreased and the false acceptance rate is much increased, compared with the contour-based algorithm. It is clear that the contour-based algorithm outperforms the region-based algorithm. The reason why the false acceptance rate of the contour-based algorithm is much less than that of the region-based algorithm is that, although there are BOIs in normal images rich in skin information, their contour-related information is different from that in the pornographic images. Compared with the region-based algorithm, the contour-based algorithm better distinguishes normal images rich in skin information from pornographic images.

8.4 Performance of Our Framework

In this section, we introduce the method for the preprocessing of Web pages and then present the accuracy of the division of Web pages into the three categories and the recognition performance of our framework.

8.4.1 Preprocessing

As the time for handling texts is much less than that for handling images, we only consider the preprocessing of images, that is, discarding the "invalid" images that are not consistent with the theme of the Web page.

The size and position of an image were used to determine its "validity." Fig. 6 shows the distributions of sizes of 8,150 images collected from the Internet, including 5,778 GIF images and 2,372 JPEG images, where a point corresponds to the size of an image. The images are marked as valid and invalid (consistent and inconsistent with the theme of the Web page) manually. For a test image, we find three points nearest to the point corresponding to the size of the image. If all the three points correspond to valid (or invalid) images, the image is treated as valid (or invalid). If the three points correspond to both valid and invalid images, we check whether the image lies at the edges of the Web page. If so, the image is treated as invalid.

8.4.2 Division of Web Pages into the Three Categories

We use Web pages collected from the Internet to check, manually, the rationality of our strategy of dividing Web pages into the three categories as described in Section 2. Table 10 shows the statistics of the categories of 1,006 collected Web pages. In the table, it is shown that 963 pages can be assigned to one of the three categories. Therefore, most of the pages drop into our division of pages. The pages that cannot be assigned to one of the three categories have the

TABLE 10
Distribution of Categories of Web Pages

Themes \ Categories	Continuous text pages	Discrete text pages	Image pages	Uncertain	Total
Art	2	57	26	8	93
Business	12	66	25	3	106
Sciences	7	58	19	6	90
Computer	9	58	23	5	95
News	14	68	30	2	114
Shopping	11	41	51	4	107
Recreation and games	8	53	45	4	110
Society	4	71	16	4	95
Health	0	72	24	2	98
Sports	5	38	50	5	98

TABLE 11
Test Results of the C4.5

Continuous text pages		Discrete text pages		Image pages	
Test	Correct	Test	Correct	Test	Correct
80	77	68	63	312	305

TABLE 12
Comparison between Individual Classifiers and Our Framework

	Continuous text pages	Discrete text pages	Image pages	Total
Continuous text classifier	97.6%	16.3%	11.1%	35.4%
Discrete text classifier	52.9%	93.3%	14.6%	69.4%
Image classifier	4.9%	74.7%	91.0%	60.4%
Our framework	96.8%	92.1%	93.4%	93.5%

characteristics of both image pages and discrete text pages, so they can be roughly divided into either the category of image pages or the category of discrete text pages.

We also evaluate the performance of our C4.5 decision tree algorithm for dividing pages into the three categories. Table 11 shows the numbers of test pages and the numbers of correctly divided pages for the three categories of pages. The accuracy rate of the division is 96.7 percent. Therefore, with a proper number of training samples and a few features, C4.5 commits very few errors.

8.4.3 Recognition Performance of Our Framework

We used the following three sets of pages to evaluate the performance of our framework: 1) **Set 1**: A set of Web pages is downloaded from the Internet and classified by hand. There are 1,500 Web pages consisting of 500 pornographic Web pages and 1,000 normal pages. Of all these pornographic pages, there are 123 continuous text pages, 288 discrete text pages, and 89 image pages. The normal pages cover 10 topics: "arts," "business," "sciences," "computer," "news," "education," "entertainment," "society," "health," and "sports." In the topics of "business" and "health," there are some legitimate sex-related contents, such as advertisements for condoms, sex-related medical materials, and sex education. 2) **Set 2**: This set of Web pages is also collected from the Internet. There are 682 Web pages, 274 of which contain dozens of pornographic images and sexually explicit words. In the 408 normal images, there are 113 legitimate sex-related pages including 69 medical pages and 44 sex education pages. 3) **Set 3**: We designed an active Web search engine (Web spider) to search for Web pages. Six Web sites are used as the roots for the search. For each root, the maximum number of searched Web pages is set to 200. A total of 1,160 Web pages were obtained. These Web pages include 411 pornographic pages and 749 normal pages, as checked by hand.

We compared the recognition rates of the three individual classifiers with those of our framework in which the results from the individual classifiers are fused. The probabilities p_1 and p_2 introduced in Section 7 are set to be 0.074 and 0.027. The results tested in page set 1 are summarized in Table 12 from which the following useful points are acquired:

TABLE 13
Test Results on Page Set 2

	Correct classification rate	False acceptance rate
Fusion on all 682 samples	91.6%	3.2%
Fusion on 113 specific samples		4.4%

- Each individual classifier obtains a favorable score on the corresponding category of test pages. However, none of them shows good performance in other categories and the overall rate is unacceptable. Therefore, it is infeasible to recognize pornographic pages using only a single-modality classifier.
- For the continuous text pages and the discrete text pages, the recognition rates of our framework are close to the best ones that the individual classifiers can obtain. For the image pages, the recognition rates of our framework are better than the best ones obtained by the individual classifiers. The total recognition rate of our framework is much higher than that of each individual classifier. Therefore, our framework makes good use of the best results from the individual classifiers.

Table 13 shows the correct classification and false acceptance rates for our framework tested on Web page set 2. For all the 682 test pages in the set, the false acceptance rate of our framework is quite low (3.2 percent) and, meanwhile, a high correct classification rate (91.6 percent) is maintained. The inclusion of pages with legitimate sex-related information may lead to more false acceptances. The table shows that, for the 113 pages of this type, the false acceptance rate is within 5 percent. Therefore, our framework can correctly classify such pages.

For Web page set 3, there are 389 pornographic pages that are correctly recognized, giving a recognition rate of 95.1 percent. There are 729 normal pages that are correctly recognized, giving a recognition rate of 97.3 percent. The reason why the recognition rates for the examples of real applications are higher than the recognition rates for the test sets 1 and 2 is that Web pages in sets 1 and 2 are selected deliberately and delicately to test the robustness of the recognition algorithms and the framework.

9 Conclusions

The existing methods for recognizing pornographic Web pages use classification of either texts or images, or simple serialization of text and image classifiers. In this paper, we propose a novel framework in which Web pages are divided, using the C4.5 decision tree algorithm, into continuous text pages, discrete text pages, and image pages. These three categories of Web pages are handled, respectively, by our continuous text classifier, our discrete text classifier, and our algorithm that fuses the results from the image classifier, and the discrete text classifier. We have the following conclusions:

1. The continuous text classifier uses the semantic features of texts to solve, to some extent, the three unsolved problems: "overblocking," "misspelled," and "wordlist."

2. The skin detection algorithm combines multidimensional histograms with the EM algorithm to speed up Gaussian mixture skin modeling without compromising the accuracy of the resulting model.
3. It is shown that the contour-based pornographic-image recognition algorithm outperforms the traditional skin-region-based algorithm.
4. The fusion algorithm implements semantic fusion of texts and images to some extent and outperforms either of the individual classifiers.
5. The framework is effective and efficient in recognizing pornographic Web pages.

Although this paper focuses on the recognition of pornographic Web pages, except for the image classifier, the continuous text classifier, the discrete text classifier, the algorithm for fusing the individual classifiers, and the algorithm for dividing Web pages are available for recognizing the Web pages dealing with other topics.

Our future work will focus on the following aspects:

- We will apply the ranking learning approach to discard texts and images inconsistent with the theme of a Web page.
- We will apply the approach for learning block importance models for Web pages [41] to fuse the information in Web pages.
- We will explore the multilevel classification strategy to classify the pornographic Web pages, excluding normal pages using a rapid classifier, and then further handling the kept pages using a more accurate classifier.
- We will use the hyperlinked relations between individual Web pages to more accurately determine whether the individual Web pages are pornographic.
- We will further investigate counterplots to antirecognition techniques.

ACKNOWLEDGMENTS

The authors appreciate Dr. Jinfeng Yang for his work on pornographic image recognition, which is the groundwork of the generalized pornographic image recognition in this paper. The authors also appreciate Professor Tieniu Tan for his valuable suggestions on the work on pornographic image recognition. This work is partly supported by the National Natural Science Foundation of China (NSFC) (Grant 60520120099 and 60672040) and the Natural Science Foundation of Beijing (Grant 4041004).

REFERENCES

[1] J.F. Yang, Z.Y. Fu, T.N. Tan, and W.M. Hu, "A Novel Approach to Detecting Adult Images," *Proc. Int'l Conf. Pattern Recognition,* vol. 4, pp. 479-482, Aug. 2004.

[2] C.D. Hunter, "Filtering the Future?: Software Filters Porn, PICS and the Internet Content Conundrum," master's thesis, Univ. of Pennsylvania, 1999.

[3] D. Keller and S.G. Verbulst, "Parental Control in a Converged Communications Environment Self-Regulation, Technical Devices and Meta-Information," *Final Report for the DVB Regulatory Group,* Univ. of Oxford, Oct. 2000.

[4] A.L. Rotta, "Report on Filtering Techniques and Approaches," Technical Report NETPROTECT: WP2:2.3v1.023, NetProtect Research Group, Oct. 2001.

[5] P.Y. Lee, S.C. Hui, and A.C.M. Fong, "Neural Networks for Web Content Filtering," *Intelligent Systems,* vol. 17, no. 5, pp. 48-57, Sept./Oct. 2002.

[6] P.Y. Lee, S.C. Hui, and A.C.M. Fong, "An Intelligent Categorization Engine for Bilingual Web Content Filtering," *IEEE Trans. Multimedia,* vol. 7, no. 6, pp. 1183-1190, 2005.

[7] R. Du, R. Safavi-Naini, and W. Susilo, "Web Filtering Using Text Classification," *Proc. IEEE Int'l Conf. Networks,* pp. 325-330, 2003.

[8] W.H. Ho and P.A. Watters, "Statistical and Structural Approaches to Filtering Internet Pornography," *Proc. IEEE Int'l Conf. System, Man and Cybernetics,* vol. 5, pp. 4792-4798, Oct. 2004.

[9] M.M. Fleck, D.A. Forsyth, and C. Bregler, "Finding Naked People," *Proc. European Conf. Computer Vision,* vol. 2, pp. 592-602, 1996.

[10] D.A. Forsyth and M.M. Fleck, "Identifying Nude Pictures," *Proc. IEEE Workshop Applications of Computer Vision,* pp. 103-108, 1996.

[11] D.A. Forsyth and M.M. Fleck, "Body Plans," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 678-683, June 1997.

[12] D.A. Forsyth and M.M. Fleck, "Finding People and Animals by Guided Assembly," *Proc. IEEE Int'l Conf. Image Processing,* vol. 3, pp. 5-8, 1997.

[13] D.A. Forsyth, S. Ioffe, and J. Haddon, "Finding Objects by Grouping Primitives," *Proc. 32nd Asilomar Conf. Signals, Systems, and Computers,* vol. 1, pp. 905-909, 1998.

[14] D.A. Forsyth and M.M. Fleck, "Automatic Detection of Human Nudes," *Int'l J. Computer Vision,* vol. 32, no. 1, pp. 63-77, 1999.

[15] S. Ioffe and D.A. Forsyth, "Finding People by Sampling," *Proc. IEEE Int'l Conf. Computer Vision,* pp. 1092-1097, 1999.

[16] S. Ioffe and D.A. Forsyth, "Learning to Find Pictures of People," *Advances in Neural Information Processing Systems 11,* M.S. Kearns, S.A. Solla, and D.A. Cohn, eds. MIT Press, 1999.

[17] S. Ioffe and D.A. Forsyth, "Probabilistic Methods for Finding People," *Int'l J. Computer Vision,* vol. 43, no. 1, pp. 45-68, 2001.

[18] J.Z. Wang, G. Wiedrhold, and O. Firschein, "System for Screening Objectionable Images Using Daubechies' Wavelets and Color Histograms," *Proc. Int'l Workshop Interactive Distributed Multimedia Systems and Telecomm. Services,* pp. 20-30, 1997.

[19] J.Z. Wang, J. Li, G. Wiederhold, and O. Firschein, "System for Screening Objectionable Images," *Computer Comm.,* vol. 21, no. 15, pp. 1355-1360, 1998.

[20] F. Jiao, W. Gao, L. Duan, and G. Cui, "Detecting Adult Image Using Multiple Features," *Proc. IEEE Int'l Conf. Info-tech and Info-net,* vol. 3, pp. 378-383, 2001.

[21] K.M. Liang, S.D. Scott, and M. Waqas, "Detecting Pornographic Images," *Proc. Asian Conf. Computer Vision,* pp. 497-502, Jan. 2004.

[22] D. Smith, R. Harvey, Y. Chan, and J. Bangham, "Classifying Web Pages by Content," *Proc. IEE European Workshop Distributed Imaging,* vol. 99/109, pp. 8/1-8/7, 1999.

[23] Y. Chan, R. Harvey, and J. Bangham, "Using Colour Features to Block Dubious Images," *Proc. European Signal Processing Conf.,* vol. 3, Sept. 2000.

[24] A. Bosson, G. Cawley, and Y. Chan, "Non-Retrieval: Blocking Pornographic Images," *Proc. Int'l Conf. Image and Video Retrieval,* pp. 50-60, 2002.

[25] W.A. Arentz and B. Olstad, "Classifying Offensive Sites Based on Image Content," *Computer Vision and Image Understanding,* vol. 94, nos. 1-3, pp. 295-310, 2004.

[26] H. Zheng, H. Liu, and M. Daoudi, "Blocking Objectionable Images: Adult Images and Harmful Symbols," *Proc. IEEE Int'l Conf. Multimedia and Expo,* vol. 2, pp. 1223-1226, June 2004.

[27] M. Hammami, Y. Chahir, and L. Chen, "WebGuard: Web Based Adult Content Detection and Filtering System," *Proc. IEEE/WIC Int'l Conf. Web Intelligence,* pp. 574-578, Oct. 2003.

[28] M. Hammami and Y. Chahir, "WebGuard: A Web Filtering Engine Combining Textual, Structural, and Visual Content-Based Analysis," *IEEE Trans. Knowledge and Data Eng.,* vol. 18, no. 2, pp. 272-284, Feb. 2006.

[29] M.J. Jones and J.M. Rehg, "Statistical Color Models with Application to Skin Detection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 274-280, June 1999.

[30] G. Eysenbach, G. Yihune, K. Lampe, P. Cross, and D. Brickley, "Quality Management, Certification and Rating of Health Information on the Net with MedCERTAIN: Using a medPICS/RDF/XML Metadata Structure for Implementing eHealth Ethics and Creating Trust Globally," *J. Medical Internet Research,* vol. 2, no. 1, suppl. 2: e1, 2000.

[31] A. Lukianiuk, "Capacity of Cellular Neural Networks as Associative Memories," *Proc. IEEE Int'l Workshop Cellular Neural Networks and Their Applications,* p. 3740, 1996.

[32] M.G. Milanova, A.C. Campilho, and M.V. Correia, "Cellular Neural Networks for Segmentation of Image Sequence," *Proc. 11th Portuguese Conf. Pattern Recognition,* pp. 49-54, 2000.

[33] T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," *Proc. 10th European Conf. Machine Learning,* pp. 137-142, 1998.

[34] A. McCallum and K. Nigam, "A Comparison of Event Models for Naive Bayes Text Categorization," *Proc. Am. Assoc. for Artificial Intelligence Nat'l Conf. Artificial Intelligence Workshop Learning for Text Categorization,* pp. 41-48, 1998.

[35] J.R. Quinlan, "Induction of Decision Trees," *Machine Learning,* vol. 1, no. 1, pp. 81-106, 1986.

[36] S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Proc. Int'l Conf. World Wide Web,* pp. 107-117, 1998.

[37] J.M. Kleinberg, "Authoritative Sources in a Hyperlinked Environment," *J. ACM,* vol. 46, no. 5, pp. 604-632, 1999.

[38] S.J. McKenna, S. Gong, and Y. Raja, "Modeling Facial Color and Identity with Gaussian Mixtures," *Pattern Recognition,* vol. 31, no. 12, pp. 1883-1892, 1998.

[39] J.C. Terrillon, M.N. Shirazi, H. Fukamachi, and S. Akamatsu, "Comparative Performance of Different Skin Chrominance Models and Chrominance Spaces for the Automatic Detection of Human Faces in Color Images," *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition,* pp. 54-61, 2000.

[40] Y. Yang, "A Comparative Study on Feature Selection in Text Categorization," *Proc. 14th Int'l Conf. Machine Learning,* pp. 412-420, 1997.

[41] R. Song, H. Liu, J. Wen, and W. Ma, "Learning Block Importance Models for Web Pages," *Proc. Int'l World Wide Web Conf.,* pp. 203-211, 2004.

[42] M.M. Kokar, J.A. Tomasik, and J. Weyman, "Data vs. Decision Fusion in the Category Theory Framework," *Proc. Fourth Int'l Conf. Information Fusion,* vol. 1, pp. TuA3-15-TuA3-20, 2001.

[43] M. Lalmas, "Dempster-Shafer's Theory of Evidence Applied to Structured Documents: Modelling Uncertainty," *Proc. 20th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval,* pp. 110-118.

[44] R. Brunelli and D. Falavigna, "Person Identification Using Multiple Cues," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 17, no. 10, pp. 955-966, Oct. 1995.

[45] http://www.ala.org/ala/washoff/WOissues/civilliberties/cipaweb/legalhistory/litigation.htm, 2006.

[46] http://www4.law.cornell.edu/uscode/html/uscode18/usc_sec_18_00002256-000-.html, 2006.

Weiming Hu received the PhD degree from the Department of Computer Science and Engineering, Zhejiang University. From April 1998 to March 2000, he was a postdoctoral research fellow with the Institute of Computer Science and Technology, Founder Research and Design Center, Peking University. Since April 2000, he has been with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. Now, he is a professor and a PhD student supervisor in the laboratory. His research interests are in visual surveillance, neural networks, filtering of Internet objectionable information, retrieval of multimedia, and understanding of Internet behaviors. He has published more than 80 papers in national and international journals, and international conferences.

Ou Wu received the BSc degree in electrical engineering from Xi'an Jiaotong University, China, in 2003. He received the MSc degree in pattern recognition from the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, in 2006. His research interests include information filtering and Web mining.

Zhouyao Chen received the BSc degree from the Department of Computer Science and Technology, Nanjing University, China, in 2004. He is currently a masteral student in the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, China. His research interests include pattern recognition and computer vision.

Zhouyu Fu received the BSc degree in information engineering from Zhejiang University, China, in 2001 and the MSc degree in pattern recognition from the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, China, in 2004. He is currently a PhD candidate in the Department of Information Engineering, Research School of Information Sciences and Engineering, Australian National University, Australia. His research interests include pattern recognition and computer vision.

Steve Maybank received the BA degree in mathematics from King's College, Cambridge, in 1976 and the PhD degree in computer science from Birkbeck College, University of London, in 1988. He joined the Pattern Recognition Group at Marconi Command and Control Systems, Frimley, in 1980 and moved to the GEC Hirst Research Centre, Wembley, in 1989. From 1993-1995, he was a Royal Society/Engineering and Physical Sciences Research Council (EPSRC) Industrial Fellow in the Department of Engineering Science at the University of Oxford. In 1995, he joined the University of Reading as a lecturer in the Department of Computer Science. In 2004, he became a professor in the School of Computer Science and Information Systems, Birkbeck College. His research interests include the geometry of multiple images, camera calibration, visual surveillance, information geometry, and the applications of statistics to computer vision. He is a senior member of the IEEE.