Neural Network Based Adult Image Classification

Conference Paper *in* Lecture Notes in Computer Science · September 2005

DOI: 10.1007/11550822_75 · Source: DBLP

4 authors, including:

Wonil Kim
Sejong University
62 PUBLICATIONS **193** CITATIONS
SEE PROFILE

Sung Wook Baik
Sejong University
187 PUBLICATIONS **1,825** CITATIONS
SEE PROFILE

Neural Network Based Adult Image Classification*

Wonil Kim[1], Han-Ku Lee[2,**], Seong Joon Yoo[1], and Sung Wook Baik[1]

[1] College of Electronics and Information Engineering,
Sejong University, Seoul,
Republic of Korea
{wikim, sjyoo, sbaik}@sejong.ac.kr
[2] School of Internet and Multimedia Engineering,
Konkuk University, Seoul,
Republic of Korea
hlee@konkuk.ac.kr

Abstract. Digital multimedia data is dramatically being increased everyday since the Internet became popular. This increment in multimedia data increases adult image contents to the Internet as well. Consequently, a large number of children are exposed to these X-rated contents. In this paper, we propose an efficient classification system that can categorize input images into adult or non-adult images. The simulation shows that this system achieved 95% of the true rate whereas it reduces the false positive rate below 3%.

1 Introduction

The Internet has made people access more information than any time before and has become the major source of information. In the other hand, it also shows the dark side. Among the millions of Web sites, there are over 500,000 web sites related to pornography and other issues that are as harmful as poison to the children [1].

We propose a neural network based classification system that can classify input images into adult or non-adult images. The visual descriptors defined by MPEG-7 are used for extracting features for a given image. These features are used as inputs for the 2 class (adult, non-adult) neural network classification system. The simulation using the Color Structure descriptor shows that the system achieved 95% of the true rate, whereas it reduces the false positive rate below 3%.

We will review several feature extraction and the image classification methods in the next section. The proposed image classification system using MPEG7 descriptors are detailed in the section 3. The simulation results and analyses of the image classifier will be discussed in the section 4. Finally the conclusion is given in section 5.

* This paper was supported by Electronics and Telecommunications Research Institute (Project No. 0801-2004-0025).
** Author for correspondence : +82-2-2049-6089.

W. Duch et al. (Eds.): ICANN 2005, LNCS 3696, pp. 481 – 486, 2005.

2 Related Works

There are two major approaches for adult image processing; one heavily depends upon a classic retrieval technique with image featuring, and another depends on an image mining technique, distinguishing adult images from normal images, using decision-boundary lines, rather than image featuring. The important point of the former is the feature extraction that is the very previous step of the image classification and that of the latter is the classifier itself, rather than the features being input to the classifier.

Most of the previous works for image rating systems does not specify these two different approaches and just presents the systems as one module. This paper discusses two different image processing techniques, and proposes a novel method that maximizes advantages of each technique and minimizes disadvantages.

2.1 Feature Extraction

In the middle of 90s, researchers and engineers became interested in detecting adult images on computers when the World Wide Web (WWW) became popular. The first algorithm to detect naked people in images was researched in [2, 8]. The main idea of the research is effectively masking skin regions using the skin filter. If skin regions that passed mask tests are matched to persons' figures, it is assumed that there are numerous naked parts in the image (e.g. the geometric filter). The algorithm detects whether or not the filtered skin color is matched to a specific body part in the whole image, rather than extracts primitive features helping effective classification.

Another approach had researched in [7]. The algorithm was based on image contents, and classified adult images by extracting skin regions and feature vectors that are useful for the image classification. The feature vector consists of color, texture, contour, placement, and relative size information for a given region. The advantages of this algorithm are that the importance of features is decided by the generic algorithm and the feature extraction is systematically (not experimentally) executed. Especially this algorithm is good to be applied to web sites rather than adult images. For instance, as a result of adopting the algorithm on 20 sites, the success rate is 89% and 11 images per second are executed.

The color histogram is one of the most common methods in the image classification. It is an effective method for the large size of data. A simple experiment based on the color histogram results in 80% of the detection rate and 8.5% of the false positive rate.

[6] proposed an efficient feature extraction system using MPEG-7 descriptor for the adult image classification. The system used three descriptors out of visual descriptors standardized by MPEG-7; edge histogram descriptor (EHD), color layout descriptor (CLD), homogeneous texture descriptor (HYD). They are effective to the adult image classification. The system also has the image database, which compares the descriptor value of a given image to that of images from the database, and retrieves 10 similar images with class information. The class of the given image is classified as the class that the majority of image is classified. Even though the results

showed that MPEG-7 descriptors can be used as very efficient features in the adult image classification, the classification method used in his paper is merely the *k*-nearest neighbor method.

The MPEG-7 descriptor is applied for extracting features in this paper. MPEG-7 has been developed by distinguished researchers for a long time, and has much effective definitions of descriptors. The visual area of MPEG-7 includes the color descriptor, the shape descriptor, the texture descriptor, and the movement descriptor. In the step of the feature extraction, the proposed system concentrates on finding descriptors to be used for the input of neural networks, rather than simply extracting values of features for descriptors.

2.2 Image Classification

As database techniques rapidly became advanced, image classification techniques that use statistical methods have improved a lot [3]. In particular, the field of data mining has been much improved, and a new field of study has appeared; image mining [4]. Thus, many research groups study and research the field of image classification via the image mining technique. The image classification can be categorized as the Neural Network, the Decision-Tree Model, and the Support Vector Machine.

Neural Network based method is the most common technique. This method concentrates on the study of decision-boundary surface telling adult images from non-adult images via the computer-based classification rule, called perceptron [10, 11]. An Artificial Neural Network (ANN) is an information processing paradigm inspired by biological nervous systems, such as the brain process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by examples. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons.

The decision tree model recursively partitions an image data space, using variables that can divide image data to most identical numbers among a number of given variables. This technique can give incredible results when characteristics and features of image data are known in advance [9].

The support vector machine technique is a brand-new image classification method. The purpose of the method is to find decision lines or surfaces distinguishing data from others like the technique using neural networks. The technique using the neural networks is just to find decision surfaces classifying the training data. But, SVM is to find decision surfaces maximizing the distance of two sets. Jiao et al. experimented on adult image classifiers using SVM [5].

We employ the Neural Network for the classification module in this paper. Inputs for the neural network are fed from the feature values extracted from MPGE-7 descriptors. Since the various descriptors can represent the specific features of a given image, the proper evaluation process should be required to choose the best one for the adult image classification.

3 Proposed System

3.1 Overall System Flow

The system uses the MPEG-7 XM program in the extraction phase to extract features from images in the database [12]. In the image database, there are about 9000 images (adult: 4500, non-adult: 4500) for training and testing. The system can use 6 descriptors; Dominant Color, Color Structure, Color Layout, Edge Histogram, Homogeneous Texture, and Region Shape. In the feature extraction module, the feature information is generated for each descriptor to extract the necessary features for the image classification. In the neural network classification step, the extracted descriptors are used for the input value and the neural networks are trained according to the 2 classes (adult and normal images).

3.2 Feature Extraction

By executing the MPEG-7 XM program, the features of training images are extracted in XML format. The feature information in XML is parsed and normalized into values between 0 and 1 with respect to values generated by each descriptor. These normalized values are used as inputs for the neural network classifier. After the phase of extracting input data used in the neural network by MPEG-7 XM, each image is attached with class information.

3.3 Classification

The neural network classifier is trained for the relation of the feature values and the corresponding class by modifying the weight values between nodes. We use the backpropagation algorithm to train the network. The classifier consists of input layer, output layer, and multiple hidden layers. The number of input nodes depends on the dimension of each descriptor, whereas the number of output nodes is two. The class information for the two output nodes is represented as (1,0) for adult images and (0,1) for normal images. In the testing process, as in the training process, the system extracts features from query images using MPEG-7 descriptors, and classifies query images using the neural network that generated by the training process.

4 Simulation and Result

In the simulation, we use MPEG-7 reference software: the e**X**perimentation **M**odel for feature extraction. The eXperimentation Model (XM) software is the simulation platform for the MPEG-7 Descriptors (Ds), Description Schemes (DSs), Coding Schemes (CSs), and Description Definition Language (DDL). Besides the normative components, the simulation platform needs some non-normative components, essentially to execute some procedural code to be executed on the data structures. The data structures and the procedural code together form the applications [12].

We simulated the 2-class image classification using 6 descriptors, such as Dominant Color, Color Structure, Color Layout, Edge Histogram, Homogeneous

Texture, and Region Shape. The classification module consists of 2 hidden layers, each with 10 nodes. The learning rate is 0.001 and the iteration number for training is 100,000.

Table 1. True / False rates of each Descriptor

Color layout

	Positive	**Negative**	**Total**
True (rate)	2107 (37.391)	2618 (46.46)	4725 (83.851)
False (rate)	207 (3.673)	703 (12.476)	910 (16.149)

Color structure

	Positive	**Negative**	**Total**
True (rate)	2551 (47.267)	2570 (47.619)	5121 (94.886)
False (rate)	143 (2.65)	133 (2.464)	276 (5.114)

Dominant color

	Positive	**Negative**	**Total**
True (rate)	1106 (29.082)	1524 (40.074)	2630 (69.156)
False (rate)	109 (2.866)	1064 (27.978)	1173 (30.844)

Edge histogram

	Positive	**Negative**	**Total**
True (rate)	2940 (57.908)	1545 (30.432)	4485 (88.34)
False (rate)	244 (4.806)	348 (6.854)	592 (11.66)

Homogeneous texture

	Positive	**Negative**	**Total**
True (rate)	493 (9.053)	2626 (48.219)	3119 (57.272)
False (rate)	108 (1.983)	2219 (40.745)	2327 (42.728)

Region shape

	Positive	**Negative**	**Total**
True (rate)	2033 (36.239)	1650 (29.411)	3683 (65.65)
False (rate)	1159 (20.66)	768 (13.69)	1927 (34.35)

Table 2. Test of Neural Network

Descriptor (dimension)	Color Layout (12)	Color Structure (256)	Dominant Color (24)	Edge Histogram (80)	Homogeneous Texture (30)	Region Shape (35)
Total	5635	5397	3803	5076	5446	5610
Correct	4725	5121	2630	4484	3119	3683
Incorrect	910	276	1173	592	2327	1927
Rate (%)	83.851	94.886	69.156	88.34	57.272	65.65

After training the network with 2600 images for each descriptor, we tested the performance of the classifier using about 5500 unused images in the testing process. Table 1 shows that the proposed image classification system performs about 95% of the true rate, whereas it reduces the false positive rate below 3 % for the Color Structure descriptor. The other descriptors, such as Edge Histogram descriptor and Color Layout descriptor, also show above par performance in both true positive and false positive rates. Table 2 shows that generally the classification using Color Layer, Color Structure, and Edge Histogram performs much better than using Homogeneous Texture, Region Shape, and Dominant Color.

5 Conclusion

This paper proposed a novel approach of applying MPEG-7 to adult image filtering systems as well as created a prototype system that can be used for the adult image classification by analyzing MPEG-7 descriptors. The visual descriptors defined by MPEG-7 are used for extracting features for a given image. These features are used as inputs for the 2 class (adult, non-adult) neural network classification system. The simulation using the Color Structure descriptor shows that the system achieved 95% of the true rate, whereas it reduces the false positive rate below 3%.

References

1. Will Archer Arentz and Bjorn Olstad, "Classifying offensive sites based on image contents", Computer Vision and Image Understanding, Vol 94, pp293-310, 2004
2. Margaret Fleck, David Forsyth, and Chris Bregler. "Finding Naked People", 1996 European Conference on Computer Vision, Vol. II, pp592-602, 1996
3. Michael J. Jones and James M. Rehg, "Statistical color models with application to skin detection", Technical Report Series, Cambridge Research Laboratory, December 1998
4. Yuna Jung, E. Hwang, Wonil Kim, Sports Image Classifier based on Bayesian Classification, Lecture Note in Artificial Intelligence 3040, 546-555, Springer, June 2004
5. Feng Jiao, Wen Giao, Lijuan Duan, and Guoqin Cui, “Detecting adult image using multiple features”, IEEE conference, Nov. 2001, pp.378 - 383 vol.3.
6. Sung-joon Yoo, “Intelligent multimedia information retrieval for identifying and rating adult images”, Lecture Note in Computer Science 3213, 165-170, Springer, June 2004
7. Mohamed Hammami, Youssef Chahir, and Liming Chen, "WebGuard: Web based adult content detection and filtering system", Proc. IEEE/WIC International Conference on Web Intelligence, pp 574-578, 2003
8. David Forsyth and Margaret Fleck, "Identifying nude pictures", IEEE Workshop on the Applications of Computer Vision 1996, pp103-108, 1996
9. David Hand, Heikki Mannila, and Padhraic Smyth, "Principles of Data Mining", MIT Press, pp343-347, 2001
10. F. Rosenblatt, "The Perceptron : A probabilistic model for information storage and organization in brain", Psychology Review 65, pp386-408
11. D.E. Rumelhart, G.E. Hinton, and R.J. Williams. "Learning representations by back-propagating errors", Nature(London), Vol. 323, pp533-536
12. http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm