World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:3, No:8, 2009

Prediction of a Human Facial Image by ANN using Image Data and its Content on Web Pages

Chutimon Thitipornvanid, and Siripun Sanguansintukul

***Abstract*—**Choosing the right metadata is a critical, as good information (metadata) attached to an image will facilitate its visibility from a pile of other images. The image's value is enhanced not only by the quality of attached metadata but also by the technique of the search. This study proposes a technique that is simple but efficient to predict a single human image from a website using the basic image data and the embedded metadata of the image's content appearing on web pages. The result is very encouraging with the prediction accuracy of 95%. This technique may become a great assist to librarians, researchers and many others for automatically and efficiently identifying a set of human images out of a greater set of images.

***Keywords*—**Metadata, Prediction, Multi-layer perceptron, Human facial image, Image mining.

I. Introduction

OVER the last few years, the interest in the digital image has grown rapidly on the World Wide Web. Visual and audio resources in the form of still pictures, graphics, audio, speech, and video play an increasingly pervasive role in our lives. Images, especially, are rich in information content. However, their content are also complex. Metadata can describe image data. Creating an image metadata is difficult, subjective, time consuming and expensive.

Typically, metadata is defined as "information about information"[1] or "information to describe other information". According to Taylor [3], metadata are referred to as a description of attributes and contents of an information package that may include descriptive information about the content, quality and condition, or characteristics of data. The objective of the metadata is to find/locate, identify, select, obtain, and navigate [23]. Metadata are used to expedite and enhance searching for resources. Metadata become important on the WWW due to the need to find useful information from the much larger available information.

Metadata in digital libraries can be divided into 3 categories: descriptive, structural, and administrative [4]. Descriptive metadata is information derived from the content of the data. Elements include: title, name, edition and publication date. Structural metadata is related to information about the structure, format and composition of the data. Administrative metadata are inherently extrinsic properties such as who, what, why, where of the object's creation and management. Metadata is not limited to documents. Any resources such as video, audio including image can be described with an appropriate metadata element set.

This information is increasingly available to the public in the electronic form. Photographs are captured and posted for various purposes. Specifically, the image of human face becomes significant on different activities such as face identification, face recognition and face tracking [5][6]. Therefore, an image searching system to enhance information retrieval on human images is expected to become of great interest.

The central focus of this study is to develop a simple but efficient technique to predict a single human image from a website using basic image data and the image content appearing on web pages. This technique may become a great assist to the librarians, researchers and many others for automatically and efficiently identifying a set of human images out of a greater set of images.

This paper is structured as follows: Section II summarizes related research works. Methodology is described in Section III. Experimental results are presented in Section IV. Finally, conclusion and discussion are given in Section V.

II. Literature Review

In the electronic environment, metadata are the value-added information of a document. Good information (metadata) attached to an image will facilitate its visibility from a pile of other images. The use of a standard metadata schema allows that information to travel between resources more effectively.

Metadata schemas, also called metamodel or metadata standards, are sets of elements with agreed definitions that meet the requirement of some communities or context. It allows information to be transferred between different databases or other resources. The most famous and widely accepted is the Dublin Core metadata [7] that was proposed in 1995 as a set of metadata standards. The semantics of Dublin Core were established among multi-disciplinary groups such as librarians, computer scientists, the museum community and related fields. The Dublin Core metadata elements, as finalized in December 1996, consists of: Title, Creator, Subject, Descriptions, Publisher, Contributor, Data, Type,

Chutimon Thitipornvanid is a graduate student in Computer Science and Information Program, Faculty of Science, Chulalongkorn University, Bangkok, 10330 Thailand (e-mail: angel_nunny@hotmail.com).

Siripun Sanguansintukul, Ph.D is a faculty in the Computer Science and Information Program, Faculty of Science, Chulalongkorn University, Bangkok, 10330, Thailand (e-mail: siripun.s@chula.ac.th).

Format, Identifier, Source, Language, Relation, Coverage, and Rights, as described in [21][22].

MARC (Machine-Readable Cataloging format)[8] is a set of standardized data structures for describing bibliographic materials that facilitates cooperative cataloging and data exchange in bibliographic information systems. http://www.loc.gov/marc/. In the 1980s (and revised in 2002), the Anglo-American Cataloguing Rules (AACR) extended the MARC standard so that it could describe music and various other kinds of "non-book" entities. MARC is often criticized for being unsuited to the modern computing environment.

XML (extensible markup language) technology such as XML Namespaces, XML Query languages, and XML database enable implementers to develop metadata schemas. XML is a simple, very flexible text format derived from SGML (ISO 8870). Typically, data that is more complicated can be transferred between databases using an XML standard. However, smaller organizations and individuals with less complicated software may need to export data manually either in the file format CSV or tab-delimited before importing it.

The Visual Resource Association (VRA)[9] is another communication standard that has developed a two-level hierarchical model for describing objects or visual works (such as paintings, sculptures, or buildings) and images of those works. A single set of metadata element, the VAR Core Categories, can be applied to the works and to the images, distinguishing characteristics of the image surrogate clearly from the characteristics of the work. The metadata schema used by the Art Museum Image Consortium (AMICO) has a similar hierarchical structure.

The latest developments in the WWW focus on the rapidly growing concept of the 'Semantic Web'. Semantic Web comes from collaborative efforts led by the W3C. The goal is to provide a common framework that will allow data to be shared and reused across various applications, across the enterprise and community boundaries. This is based on the idea that existing methods of finding, sharing and combining information on the web can be extended by enabling computers to understand the meaning (semantics) within web resources. For example, Google's image search uses this approach, analyzing the contextual information around images to assess their relevance in searching.

Finally, the value of an image is enhanced not only by the quality of its attached metadata, but also by the way in which individual stock library search engines interact with that metadata, and use sophisticated search algorithms to refine their search functionality. An increased use of metadata standards will support cross-searching and the sharing of images and metadata records between different systems and collections – the so-called 'metasearch'.

Various kind of methods such as template matching [10] and Eigenfaces [11] have been proposed to detect human facial images. However, most are researches concerning computer vision with object recognition [12][13][14], typically very complex and time consuming.

The methodology employed here is simple and believed to be very effective. It utilizes information from both the image itself and the contents surrounding the image on the web pages. All information obtained is then used to train the artificial neural network so that the network learns to predict whether it is a human facial image.

III. Methodology

This study presents a simple but effective technique, utilizing the both low and high level image information; content on web pages are used to increase the accuracy rate on the prediction on a single human facial image. The system is particularly valuable for a human image search.

The procedure in the experimental study is illustrated as in Fig. 1.

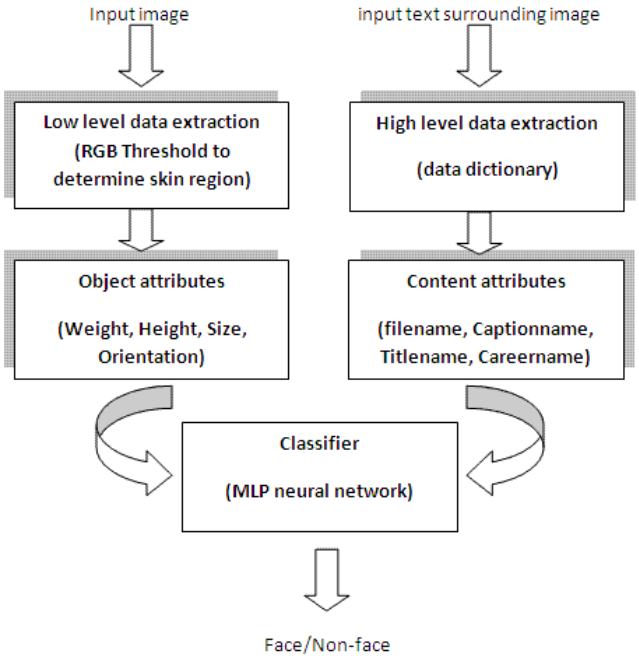

Fig. 1 The experimental procedure

The experimental procedure can be divided into these following steps:

A. Data extraction
   1.1 Low level data extraction
   1.2 High level data extraction
B. Training the network
C. Testing the performance of model from the network.

A. Data Extraction

An automated program is developed to gather images and image descriptions from CNN websites during the month of Jan-March 2009. There are altogether 400 images. These collected images will be further employed in data extraction processes as follows:

1) Low level data extraction

Each image in the form of compressed image file (GIF) format is fed into an automated image processing system [15].

If a human face is detected, the program will identify the facial skin, using a RGB color model, such as color and texture.

The objective is segmenting the skin tone color pixel from the background. The thresholds based on RGB color model are chosen as shown in the equation (1).

$$g_{rgb}(x,y)=\begin{cases}(r,g,b);if\ (R(x,y)\in[120,255],G(x,y)\in[100,225],B(x,y)\\ \in[100,225])\wedge(R(x,y)\neq G(x,y)\neq B(x,y))\\ (0,0,0)\ ;\text{otherwise}\end{cases} \quad (1)$$

The image information (pixels) such as height, width, size, and orientation of an image objects are then recorded. Therefore, there are 4 attributes obtained from this low-level extraction. Fig. 2 shows a sample image and its detected attributes from the system. (The detected object is in the face box)

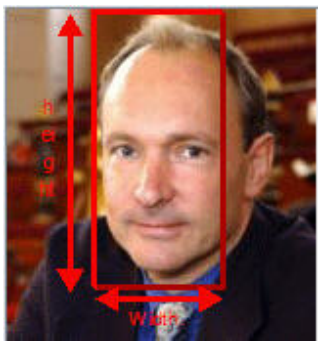

Fig. 2 A Sample of human facial image

2) High level data extraction

Information related to the image on web pages is obtained. Then, the obtained information (words) attributes are analyzed using lexitron for Nectec[20], a thai-English dictionary. Four attributes information obtained are:

1. file name : define an image file name.
2. caption name : define a table caption
3. alt name : define alternate text for an image. It is an author-defined text.
4. role and position : define role and position such as Secretary, Commentary, Prime minister, CEO. These words, typically, start with capital letters and acquire from title and caption tags.

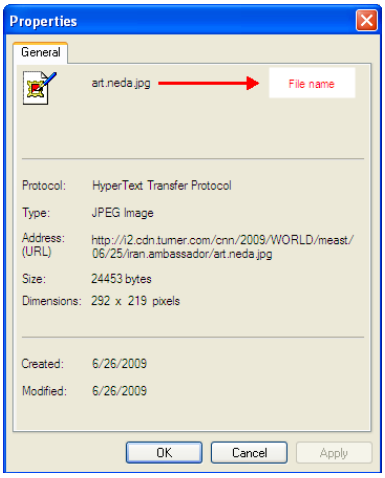

Fig. 3 An example of file name

To process the file name, the system will keep only neda.jpg from the CNN website. Then the word 'neda' is compared with the database corpus of the dictionary. If the compared name is a proper name, set the value to 1, Otherwise, the value is set to 0.

To simplified the process of caption name, words with capital letters will be examined. For example, caption "Neda, Agha-Soltan, Tehran, Saturday". Each word will be analyzed. If the word is determined as a proper name, the value is to 1. Otherwise, the value is set to 0.

Fig. 4 An example of caption name

Again, to simplify the process of title name, only words with capital letters will be evaluated. The followings display an example of tag title name from view source code in the HTML file.

<title>Iranian envoy: CIA involved in Neda's shooting? - CNN.com</title>

Words: CIA and Neda will be evaluated using lexitron [20] whether these words are proper name. If so, the value is set to 1. Otherwise, the value is set to 0.

To process role and position, words obtained from both caption and title which are 6 words in this case (2 from title, 4 from caption) are investigated again. Each word is analyzed

with the dictionary program. Examples of these words are Secretary, Commentary, Prime Minister, Singer.

Therefore, 4 additional attributes are obtained from the image content. Finally, there are 8 input attributes obtained from both the low and high level data extraction.

B. Training the Network

The artificial neural network (ANN), or neural network in short, is inspired by simulating the function of human brain. A neural network can be used to represent a nonlinear mapping between input and output vectors. Neural networks are among the more popular signal-processing technologies. In engineering, neural networks serve two important functions: as pattern classifiers and as nonlinear adaptive filters. A general network consists of a layered architecture, an input layer, one or more hidden layers and an output layer.

Fig. 5 shows a typical architecture of a multilayer perceptron network. A Multi-layers Perception (MLP) is a particular kind of artificial neural network. The MLP is used extensively to solve a number of different problems, including pattern recognition and interpolation [16]. Each layer is composed of neurons, which are interconnected with each other by weights. In each neuron, a specific mathematical function called activation function accepts input from previous layer and generates output for the next layer. In the experiment, utilized activation function is the Hyperbolic tangent sigmoid transfer function [17]. The MLP is trained using a standard back-propagation algorithm[18].

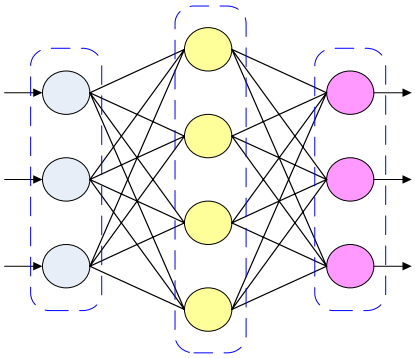

Fig. 5 A typical Multilayer Perceptron ANNs Architecture

The multiplayer perceptron (MLP) is trained using a well-known machine learning suite: WEKA[2].

IV. Experimental Results

In the experiment, the input data consist of approximately 400 images and these images are automatically collected from the CNN website. The images are in the format of GIF files. The ratio of the training set and testing set is 60:40. Therefore, 400 images are divided into 240 images for training and 160 images for testing the performance of the network.

Choosing the number of hidden units is an important factor. There are several publications discussing the number of hidden units, such as Elisseeff, et al [19]. The equation to estimate the number of hidden unit (H) is given as:

$$H \geq \frac{n-m}{m(k+2)} \tag{3}$$

k is the number of inputs equal to the number of attributes. N is the sample size (400). In order to light over fit the data, there must be fewer than m cases for each parameter. Normally, m set to 10. The value of hidden units can be calculated by replacing each parameter with its corresponding value using the above formula as following:

$$H \geq \frac{400-10}{10(8+2)} \approx 4$$

The number of hidden units initially are 4. However, hidden units are varied to get the optimal results. The architecture of the network in this study finally becomes 8 input nodes, 8 hidden nodes and 2 output nodes (8:8:2)

The learning rate and momentum are set to 0.2 and 0.3, respectively.

Table I shows the confusion matrix of the predictive results of the network. It can be seen that there are 100 true positive (n_{TP}) images, 52 true negative (n_{TN}) images, 2 false positive (n_{FP}) images and 6 false negative (n_{FN}).

TABLE I
THE FACE CLASSIFICATION RESULT

PREDICTIION OUTCOME	ACTUAL	
	FACIAL IMAGE	NON FACIAL IMAGE
FACIAL IMAGE	100	2
NON FACIAL IMAGE	6	52

In addition, the performance (accuracy) of the system is calculated by the following formula:

$$ACC = \frac{n_{TP} + n_{TN}}{n_{TP} + n_{FP} + n_{TN} + n_{FN}} \tag{4}$$

$$ACC = \frac{100 + 52}{100 + 2 + 52 + 6} = 95\%$$

V. Discussion and Conclusion

A multi-layer perceptron is trained to predict a single facial human image using the image data and the embedding metadata surrounding the image as the input to the network. The prediction result is very encouraging with the accuracy up to 95%.

However, several challenge issues need be addressed such as 1) changing the RGB color model, which is employed in the lower-level data extraction, to a different color model may improve the performance of the system. 2) extending with other photographic embedding metadata to see whether it builds a better metadata. 3) a more controlled vocabulary for describing specifics of human facial images may improve the

ability to discover image resources.

Finally, there should be a balance between what users ask for and what the metadata can support. These developments have heightened the need for effective image retrieval techniques. If the metadata are consistent, it will allow the search engine to generate good hit lists.

ACKNOWLEDGMENT

We would like to thank Worasit Chuchaiwattana, Ph.D for providing the program to extract the information from the image.

REFERENCES

[1] Heery R, Powell A, Day M. "Metadata. Library Information Technology Centre," South Bank University, London, 1997(Library & Information Briefings,85)
[2] Ian H. Witten and Eibe Frank. Data Mining: Practical Machine Learning Tools and Techniqueues. Morgan Kaufmann Publishers, San Francisco, second edition, 2005
[3] Robert S. Taylor, "Information use environments," In progress in Communication Science, ed. B. Dervin and M.J. Voigt. Norwood, New Jersey: Alblex Publishing, 1991.
[4] Arlene G. Taylor and Daniel N.Joudrey, "The organization of information," 3rd ed. Westport, Conn.: Libraries Unlimited, 2009, pp. 96-103.
[5] R. Chellappa, C.L.Wilson, and S.Sirohey, "Human and machine recognition of faces: a survey", IEEE Trans. Digital Object Identifier, Vol. 83, pp. 705-741, May 1995.
[6] Ming Hsuan Yang, David J. Kriegman, and Narendra Ahuja, "Detecting Face in Images: A Survey," IEEE Trans. Pattern Analysis and Machine intelligence, Vol. 24, pp. 34-58, Jan. 2002.
[7] Stuart L. Weibel, "The State of the Dublin Core Metadata Initiative 1999", in D-Lib Magazine April 1999, Vol. 5.
[8] Betty Furrie, "Understanding MARC Bibliographic: Machine-Readable Cataloging," 7th ed. Cataloging Distribution Service, pp. 1-29, Apr. 2003.
[9] Ben Kessler, "Encoding Works and Images: The Story Behind VRA core 4.0," Feature Articles VRA Bulletin, Vol. 34, No.1, Spring. 2007.
[10] G. Antoniol, G. Casazza, M.D. Penta, and R. Fiutem, "Object-oriented design pattern recovery," Journal of Systems and Software, pp. 181-196, 2001.
[11] A. Pentland and M. Turk, "Face recognition using eigenfaces," IEEE Computer Vision and Pattern Recognition, vol. 4, pp. 586-591, Jun 1991.
[12] M-H Yang, D. Kriegman and N. Ahuja, "Detecting Face in Images : A Survey ", IEEE Trans, Pattern Analysis and Machine Intelligence, vol. 24, no. 1, pp. 34-58, Jan. 2002
[13] Kah Kay Sung and Tomaso Poggio, "Example-based learning for view-based human face detection," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, pp. 39-51, Jan. 1998.
[14] Paul Viola and Michael J. Jones, "Robust real-time object detection," International Journal of Computer Vision, Feb. 2001.
[15] Choochaiwattana W., Niranartlumphong W., and Spring M.B., "Web image classification algorithm: a heuristic rule-based approach," ITA07, The Second International Conference on Internet Technologies, Wrexham, North East Wales, UK September 4-7, 2007.
[16] F. Soulie, E. Viennet, and B. Lamy, "Multi-Modular NeuralNetwork Architectures: Pattern Recognition Applications inoptical character Recognition and Human Face Recognition," Int'l J. Pattern Recognition and Artificial Intelligence, vol. 7, no. 4, pp. 721-755, 1993.
[17] S. P. Bingulac, "On the compatibility of adaptive controllers (Published Conference Proceedings style)," in *Proc. 4th Annu. Allerton Conf. Circuits and Systems Theory*, New York, 1994, pp. 8–16.
[18] Henry A. Rowley, Shumeet Baluja, Takeo Kanade, "Neural Network-Based Face Detection," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 1, pp. 23-38, Jan. 1998.
[19] G. R. Faulhaber, "Design of service systems with priority reservation," in *Conf. Rec. 1995 IEEE Int. Conf. Communications*, pp. 3–8.
[20] Available: http://lexitron.nectec.or.th/2009_1/
[21] Stuart Weibel, "The State of the Dublin Core Metadata Initiative," vol. 5. D-Lib Magazine, Apr. 1999. Available: http://www.dlib.org/dlib/april99/04weibel.html
[22] Renato Iannella, "Metadata: Enabling the Internet," Available: http://www.ifla.org.sg/documents/libraries/cataloging/metadata/ianr1.pdf
[23] International Federation of Library Association and Institutions (IFLA), Guidance on the structure, content, and application of descriptive metadata records for digital resource and collections". Cataloguing working group the use of Metadata schemes, 2005.