Machine Annotation for Digital Imagery of Historical Materials Using the ALIP System

James Z. Wang
School of Information Sciences and Technology
The Pennsylvania State University

Jia Li
Department of Statistics
The Pennsylvania State University

Ching-chih Chen
Graduate School of Library and Information Science
Simmons College

Abstract

Manual annotating digital imagery of historical materials is a labor-intensive task for many historians. In this paper, we introduce the application of the ALIP (Automatic Linguistic Indexing of Pictures) system, developed at The Pennsylvania State University, to the problem of machine assisted annotation of these images. The ALIP system learns the expertise of a human annotator from a small collection of representative images. The learned knowledge about the domain-specific concepts is stored as a dictionary of statistical models in a computer-based knowledge base. When a new image is presented to ALIP, the system computes the statistical likelihood of the image resembling each of the learned statistical models and the best few are further studied for the purpose of annotating the image with keywords. The Penn State research team applied their ALIP system to the EMPEROR images base and metadata created by C.-c. Chen. Promising results have been obtained.

1. Introduction

Annotating digital imagery of historical materials is labor-intensive. Typically, a well-trained human annotator must go through individual images and type in keywords or linguistic descriptions. As the image databases grow larger and larger, it is becoming prohibitively expensive to annotate these images manually. Is it possible for computers to learn the expertise from historians and use the learned knowledge to annotate images automatically and linguistically? This is the question we attempt to address.

Figure 1. Sample images selected from the EMPEROR database.

Digital images of the same type of historical objects often have somewhat similar looks. We hypothesize that a computer program should be able to learn the human expertise from some sample annotations. For example, the First Emperor of China, Qin Shi Huang Di (259-210 B.C.), has his mausoleums surrounded by more than seven thousand life-sized terra-cotta warriors and horses. Each warrior is unique, being modeled after an actual guard-of-honor. It should be possible to train computers with the concept of "Terracotta Warriors and Horses" with a few examples and let computers annotate other such images. C.-c. Chen of Simmons College, a co-author of this paper, created extensive documentary using the historical materials related to the First Emperor of China and created *The First Emperor of China*'s interactive videodisc as well as multimedia CD-ROM, both published by the Voyager Company in 1991 and 1993. The extensive image collection was further expanded as a part of her Chinese Memory Net image database since 1999 [Chen, 2001]. Figure 1 shows some samples of her image database. She also created extensive metadata including keyword and descriptive annotations to thousands of these photos manually. Her effort has made it possible for computer scientists to study the suitability of applying the state-of-the-art machine learning techniques to images of historical materials.

While content-based image retrieval techniques make it possible to search for similar images using feature similarity analysis [Smeulders, 2000] [Wang, 2001], machine annotation of images is generally considered impossible because of the great difficulties in converting the structure and content of an image into linguistic terms.

The Penn State research team has been developing the Automatic Linguistic Indexing of Pictures (ALIP) system since 2000 [Wang and Li, 2002] to learn objects and concepts through image-based training. The system was inspired by the fact a human being can recognize objects and concepts by matching a visual scene with the knowledge structure stored in the brain. For example, a 3-year old child is able to recognize a number of concepts or objects. The ALIP system builds a knowledge base about different concepts automatically from training images. Statistical models are created about individual concepts by analyzing a set of features extracted using wavelet transform [Wang, 2001]. A dictionary of these models is stored in the memory of the computer system and used in the recognition process. The team has conducted large-scale learning experiments using general-purpose photographic images representing 600 different concepts. It has been demonstrated that the ALIP system with 2-D multiresolution hidden Markov models (2-D MHMM) [Li, 2000] is capable of annotating new images with keywords after being trained with these concepts [Li and Wang, 2003].

With the funding from the US National Science Foundation, the research team has been studying the use of ALIP for annotating digital imagery of historical materials since August 2002. They are attempting to determine whether it is possible for ALIP to learn domain-specific knowledge from the human annotations of historical material images. The EMPEROR image database is suitable for this task because of both the high quality of the images and the comprehensiveness of the metadata descriptions. In the rest of the paper, we will present the training and the machine annotation experiments we have conducted. The work is still on-going. We expect to report more comprehensive results in the near future.

2. Training ALIP for the EMPEROR Collection

To annotate images automatically, we first need to train the ALIP system with different concepts. In our initial experiment, five different concepts in the EMPEROR collection are selected to demonstrate the feasibility of the approach. These concepts are: (1) Terracotta Warriors and Horses, (2) The Great Wall, (3) Roof Tile-End, (4) Terracotta Warrior – Head, and (5) Afang Palace – Painting.

Figures 2 and 3 show the images used to train two of the five concepts. For the concept of "Terracotta Warriors and Horses", a total of only eight images are used to train ALIP. For the concept of "Roof Tile-End", a total of four images are used. Just like training human beings, the more complex the concept, the more training is typically required. One of the key advantages of the ALIP approach is that training images for a concept are not required to all visually similar. The system is highly scalable because the training of one concept does not involve images related to other concepts. If more images are added to the training collection for a given concept, only that concept needs to be retrained.

Because of the small number of training images per concept, only a couple of minutes of CPU time are required to train a concept on a Pentium PC running LINUX operating system. The training process is parallelizable because the training of one concept requires only images related to that concept. For our prior experiments with 600 concepts, we used a cluster of many Pentium computers.

Figure 2. Eight images are used to train the ALIP system for the concept "Terracotta Warriors and Horses".

Figure 3. Four images are used to train the ALIP system for the concept "Roof Tile-End".

3. Machine Annotation Experiments

To validate the effectiveness of the ALIP training, we tested ALIP with other EMPEROR images related to the five trained concepts. In another word, we give the ALIP system an examination on how it has learned during such a short period of time.

It takes the computer only a few seconds to compute the statistical likelihoods of an image to all five learned concepts and sort the results. The concept with the highest likelihood is used to annotate the image. The experimental results are summarized as follows:

(1) Terracotta Warriors and Horses: A total of 52 images were tested. Only one image was mistakenly annotated as "The Great Wall". The accuracy for this concept is 98%.

(2) The Great Wall: A total of 65 images were tested. Only one image was mistakenly annotated as "Terracotta Warriors and Horses". The accuracy for this concept is 98%.

(3) Roof Tile-End: A total of 28 images were tested. Three images were mistakenly annotated as "The Great Wall". Two images were marked as "Terracotta Warriors and Horses". The accuracy for this concept is 82%.

(4) Terracotta Warrior – Head: A total of 57 images were tested. Two images were mistakenly annotated as "The Great Wall". The accuracy for this concept is 96%.

(5) Afang Palace – Painting: A total of 33 images were tested. Six images were mistakenly annotated as "The Great Wall". The accuracy for this concept is 82%.

Figure 4 shows those images mistakenly marked by the ALIP system. In our experiment, we trained each concept with only four to eight images. The experimental results are reasonable, considering that only a small amount of training is used. We expect the performance to improve if more images are used for training. In the future, we will conduct larger scale experiments so that more concepts from the EMPEROR collection can be trained.

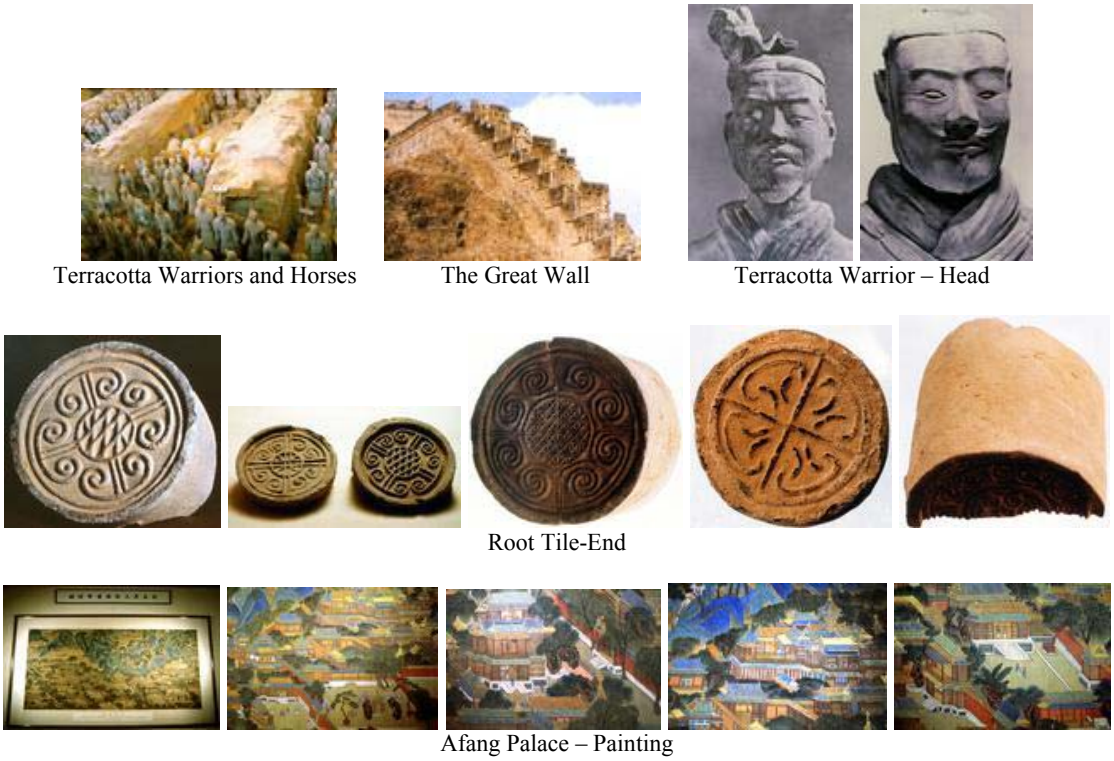

Figure 4. Images mistakenly marked by the ALIP system. The correct annotations are shown.

4. Conclusions and Future Work

In this paper, we have demonstrated that it is possible to train the Penn State ALIP system for annotating digital imagery of historical materials. The EMPEROR image collection is used in the experiments. Promising results have been obtained. In the future, we will conduct large-scale learning and annotation experiments. On the technology side, we are refining the system with improved models.

Acknowledgments

The material is based upon work supported by the US National Science Foundation under grant no. IIS-0219272, the Pennsylvania State University, the PNC Foundation, and Sun Microsystems under grant EDUD-7824-010456-US. EMPEROR images collection and metadata provided by C.-c. Chen is a part of her Chinese Memory Net support by NSF/IDLP under grant no. IIS-9905833. Conversations with Michael Lesk, Stephen Griffin, and members of the DELOS-NSF Working Group on Digital Imagery for Significant Cultural and Historical Materials have been very helpful.

References

[Chen, 2001] C.-c. Chen, "Chinese Memory Net (CMNet): A model for collaborative global digital library development," *Global Digital Library in the New Millennium: Fertile Ground for Distributed Cross-Disciplinary Collaboration.* C.-c. Chen, ed., Beijing: Tsinghua University Press, pp. 21-32, 2001.

[Li and Wang, 2003] J. Li and J. Z. Wang, ``Automatic linguistic indexing of pictures by a statistical modeling approach, '' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, 14 pp., 2003.

[Li, 2000] J. Li and R. M. Gray, *Image Segmentation and Compression Using Hidden Markov Models*, Kluwer Academic Publishers, Dordrecht, 2000.

[Smeulders, 2000] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Analysis And Machine Intelligence*, vol. 22, no. 12, pp. 1349-1380, Dec. 2000.

[Wang and Li, 2002] J. Z. Wang and J. Li, "Learning-based linguistic indexing of pictures with 2-D MHMMs," *Proc. ACM Multimedia*, pp. 436-445, Juan Les Pins, France, ACM, December 2002.

[Wang, 2001] J. Z. Wang, *Integrated Region-Based Image Retrieval*, Kluwer Academic Publishers, 190 pages, Dordrecht, 2001.