EXTRACTING SEMANTIC INFORMATION FROM ART IMAGES *

Elena Šikudová,[1] Marios A. Gavrielides,[2] Ioannis Pitas[2]
[1] *Dept. of appl. CS, Comenius University, Bratislava, Slovakia*
[2] *AIIA Laboratory, Aristotle University of Thessaloniki, Greece*

Abstract *This paper presents a method for automatic annotation of portraits in art image databases and discusses the extraction of semantic information from portraits. The proposed method segments images into candidate regions and fits an ellipse and a bounding box to them. Their extracted features serve as input to a neural network, which is trained to distinguish between face and non-face regions. Paintings containing face regions are classified as portraits. The method evaluation is done on a set of 188 digital paintings using ROC curves as performance measures. The results show that the method is very efficient in locating the face regions and in the recognition of portrait paintings. The performance of the algorithm is encouraging for its further development, which includes the extraction of portrait-specific semantic information.*

Introduction

Modern electronic image libraries, including Bridgeman Art Library and Getty Images, allow the users to browse and search through fine art image databases. Locating certain images in such a database relies on textual image annotations, previously embedded in the database. Typically, the image annotation is done manually by experts at the archival institutions. The labor-intensive process of manually annotating thousands of images calls for automatic or semi-automatic annotation tools.

As a first step towards an automatic mechanism for image annotation, we have developed a method that identifies digitized fine art images as portraits or non-portraits, and which extracts information about the number of faces, their location, size and spatial arrangement in the image. This information would allow users an access to more sophisticated queries. We assume that a portrait shows one or more persons with their faces mainly in frontal view, and in the foreground of the image, with-

*This work has been supported by the European Union project RTN1-1999-00177 MOUMIR (Methods for Unified Multimedia Information Retrieval) and by VEGA grant Virtual Environments for WWW No. 1/0174/03. The authors would like to thank the Bridgeman Art Library for providing the painting database for this study.

out any occlusion. Since the faces can be of different sizes, we use the assumption that the face in a portrait covers at least 1% of the painting area.

Related work for portrait image annotation can be found in [1] and [2]. However both papers address only annotation of photographs and not that of fine art images. Our method can be also seen as an application of face detection methods into the fine art domain. Recently an extensive survey on existing face detection methods appeared in [3]. Many of the described face detection systems use assumptions, which prevent them from being used in automatic annotation of fine art images. Among these assumptions are the use of gray scale images only, fixed image size, non-complex background and the use of "mug-shot" images. Our study focuses on the digitized paintings, uses the color information of images, can deal with images of arbitrary size, complex background, faces of different sizes and multiple face occurrence.

This paper is organized as follows. Section 1 describes each stage of the proposed algorithm. Section 2 presents the achieved results and the evaluation of the proposed method. The final section contains discussion on the results and elaborates on an indexing scheme under development that would allow the extraction of semantic information related to a given portrait.

1. Method

The individual steps of the proposed multistage algorithm for the automatic recognition and annotation of portraits in art images databases are as follows.

First, a painting is segmented into candidate regions. An elliptical class boundary based on Gaussian distribution of colors identifies the skin colored pixels. Foreground pixels inside this boundary are treated as possible skin pixels. The resulting connected regions are further divided, based on the edge boundary presence. Then an ellipse and a bounding box are fitted to the segmented face candidate regions, from which a set of feature vectors is extracted. The vectors are subsequently fed to an artificial neural network that is trained to distinguish between face and non-face regions. At the end, images containing face regions are classified as portrait images. The method was evaluated both for its ability to classify images correctly and for its accuracy in locating the face region.

The intermediate results of each step are illustrated in the Figure 1b-d. The original *Portrait of M. A. Bek*, by Karl Pavlovich Bryullov (©Bridgeman Art Library) is shown in Figure 1a.

1.1 Region segmentation

The HSV color model, where the H, S and V components describe color hue, saturation and intensity respectively, was chosen for skin color classification. The values of S and V lie in the range $[0, 1]$, whereas the values of H are circular in range $[0, 2\pi]$ with $H = 0$ and $H = 2\pi$ defining the same hue (red). In order to circumvent the need for using the angular statistics [4], we benefited from the fact that the skin color cluster is localized in proximity of the red hue of HS space. The hue coordinate H was therefore transformed into H' in a non-circular interval $[-\pi, \pi]$.

To classify the pixels, an elliptical skin color boundary was used, based on a two-dimensional Gaussian probability density function (pdf). The mean values $\overline{H'}, \overline{S}$ and covariance matrix parameters were estimated using pixels from manually extracted face regions in the training set. Only pixels with pdf value greater than an empirically determined threshold were labelled as skin colored. The threshold was set after analyzing the correct classification rates of the skin and non-skin pixels in the training set. The result of color-based pixel classification for the portrait example is shown in Figure 1b.

The presence of an edge was used to divide connected regions of the foreground skin colored pixels. This step was particularly useful when splitting the face from the neck area. The edges in the original intensity image were exploited. Also, a size limit was applied to remove small regions. After analyzing the portraits in the available database, the threshold was set to 1% of the image size. The analysis showed also that the faces were not found at the sides and the bottom of an image. Therefore pixels lying in the margin areas were removed.

1.2 Feature extraction

In order to classify face candidate regions identified in the previous section, a feature vector was extracted from a bounding box and an ellipse fitted to each region. An axis-aligned bounding box of the region was used. The extracted features are the width and the height of the bounding box relative to the image width and height respectively.

(a) (b) (c) (d)

Figure 1. Illustration of individual steps of the method. a) Original image. b) After color pixel segmentation. c) Ellipses fitted to the segmented candidate regions. d) Final output.

The face region features associated with the fitted ellipse are defined as follows. The orientation of the ellipse determines the angle between the x axis and the major axis of the ellipse. The aspect ratio of the axes is the ratio of the major axis to the minor axis. The face region relative horizontal and vertical placement is the position of the ellipse center relative to the horizontal and vertical image size. The region/ellipse overlap feature determines what percentage of the best-fit ellipse is covered by the region. The left-out percentage feature determines what percentage of the region is not covered by the best-fit ellipse. Ellipses fitted to the resulting face candidate regions are shown in Figure 1c.

1.3 Region classification

The features were extracted from overall 188 images, divided into two disjoint subsets. The training set contained 88 images (38 portraits, 50 non-portrait pictures), and the testing set contained 100 images (50 portraits, 50 non-portrait pictures). The ground truth location of the faces in images was manually determined and the extracted regions were labelled with target values 1 (face) and 0 (non-face). The criterion for labelling the region with value 1 was that the ratio of the overlap between the true and segmented face regions and union of these two regions was greater than 0.5. The segmentation of the images resulted in 38 face regions and 322 non-face regions in the training set and 50 face regions and 280 non-face regions in the testing set.

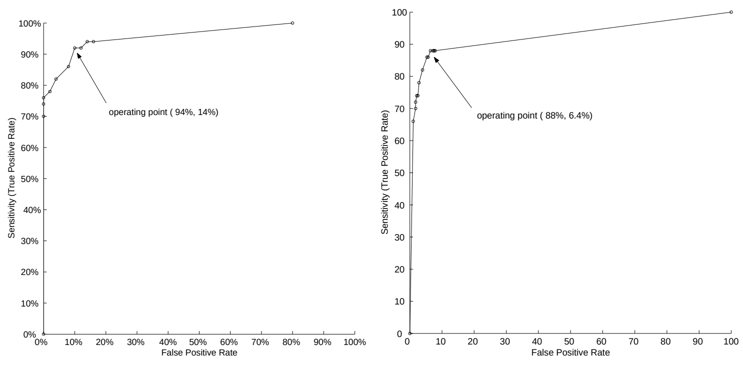
A three-layer back-propagation artificial neural network (ANN) was trained to classify the candidate regions as a face or non-face using as input the values of the extracted ellipse and bounding box features. The ANN consisted of the input layer, one hidden layer containing up to 18 nodes and the output layer with one node. The network output is a value in the range $[0, 1]$. Network parameters such as the number of hidden nodes, momentum, learning rate and number of iterations were determined empirically from the evaluation of the ANN performance on the training set. The ANN was implemented using the Neural Network Toolbox of the MATLAB software [5].

As a consistency test, the ANN was tested on the training set in order to determine its ability to learn from examples. In this test the ANN was able to classify correctly all face regions without any falsely classified non-face regions. After training, the weights of the neural network were fixed and the network was applied to the feature vectors acquired from the testing set.

Regions with ANN output value less than a threshold were removed from the final output, as it can be seen in Figure 1d for the example portrait.

2. Results

The algorithm was evaluated on the testing set after being trained exclusively on the training set, in order to get an assessment of the performance of the algorithm when applied to an unknown set of cases. The performance of the proposed method was evaluated using two different ways. In the *per-image basis* evaluation, the method was tested for its ability to classify an image as a portrait or a non-portrait. The criterion for classifying an image as a portrait was the presence of a detected face, defined as a region with a ANN output higher than a predefined threshold. In the *per-region basis* evaluation, the method was tested for its ability to correctly classify segmented regions as either face or non-face regions. By correctly classifying face regions, the output of the algorithm can be used in a subsequent face analysis step to extract semantic information from the images. The Receiver Operating Characteristic



(a) (b)

Figure 2. ROC curve describing the results in the a) per-image and b) per-region evaluation.

(ROC) methodology was used to describe the performance of our algorithm [6]. The points on a ROC curve were acquired by comparing the ANN output with a threshold ranging from 0 to 1. Each point of the curve is a pair of *sensitivity* and *false positive rate* (FPR). Sensitivity is defined as the number of correctly classified *cases* over the total number of *cases* in the set, FPR as the number of falsely classified *cases* over the total number of *cases* in the set, where *cases* refer to portraits in the per-image evaluation and to face regions in the per-region evaluation.

The resulting ROC curve for the per-image evaluation is shown in Figure 2a. As it can be seen, the method can achieve the sensitivity level of 94% at the false positive rate of 14%. The ROC curve describing the per-region performance is shown in Figure 2b. It can be seen that the sensitivity rate of 88% for detecting a face region can be reached at the false positive rate of 6.4%.

3. Discussion and conclusion

The results of both evaluations stated in the previous section, proved the potential of the proposed algorithm to correctly identify portrait images within a general database. Besides automatic annotation, the method could be used as part of an image retrieval system that would allow for queries specific to fine art images.

In order to extract semantic information from portraits, we are currently developing a method that will follow the detection of faces in an image by generating a description of the number, the size and the relative location of faces within the image. The mutual arrangement of the faces in the descriptor can be expressed using polar coordinates, stating the distance and the angle between faces. For each face the center of the fitted ellipse (from the face detection procedure) is the center of local coordinate system. The faces below the "origin" face form a positive angle, faces above form a negative angle, as illustrated in Figure 3. The proposed method could be improved by further analyzing the segmented face regions in order to locate facial features. The presence of eyes, nostrils and mouth could verify the classification of a face and reduce the number of false positives. Nevertheless, the performance of the algorithm is encouraging for its further development, which includes the extraction of portrait-specific semantic information.

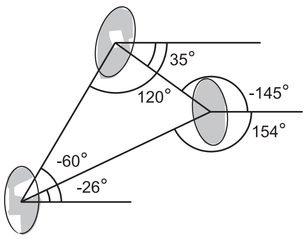


Figure 3. Example of mutual spatial layout of faces.

References

[1] Saber, Eli, Tekalp, A. Murat, Eschbach, Reiner, and Knox, Keith (1996). Automatic Image Annotation Using Adaptive Color Classification. *Graphical Models and Image Processing*, 58(2):115–126.

[2] Gevers, Theo, Aldershof, Frank, and Smeulders, A.W.M. (2000). Classification of Images on Internet by Visual and Textual Information. *IST/SPIE Electronic Imaging, Internet Imaging*, 3964:16–27.

[3] Yang, Ming-Hsuan, Kriegman, David J., and Ahuja, Narendra (2002). Detecting Faces in Images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58.

[4] Mardia, K. V. (1972). *Statistics of Dicrectional Data.* Academic Press.

[5] Demuth, H. and Beale, M. (1993) *Neural Network Toolbox: For use with MATLAB: User's Guide.* The Mathworks.

[6] Metz, C. E. (1978). Basic Principles of ROC Analysis. *Seminars in Nuclear Medicine*, 8:283–298.