Painter Identification Using Local Features and Naive Bayes

Daniel Keren
Department of Computer Science
University of Haifa
Haifa 31905, Israel
dkeren@cs.haifa.ac.il

Abstract

The goal of this paper is to offer a framework for image classification "by type". For example, one may want to classify an image of a certain office as man-made – as opposed to outdoor – scene, even if no image of a similar office exists in the training set. This is accomplished by using local features, and using the naive Bayes classifier.

The application presented here is classification of paintings; after the system is presented with a sample of paintings of various artists, it tries to determine who was the painter who painted it. The result is local – each small image block is assigned a painter, and a majority vote determines the painter. The results are roughly visually consistent with human perception of various artists' style.

1. Introduction

One of the visual tasks which human perform extremely well may be described as "recognition by type". For example, a person can quite successfully determines the identity of an artist who drew a certain picture, given that he/she is familiar with other pictures made by the artist, and that two painters with very similar styles are not present (still, in that case, a person will be able to recognize the school of the painting - cubist, expressionist etc.).

The recognition of "style" does not use gray level or color similarity, nor high-level features (such as faces, eyes etc.), which excludes using many methods which are successful for other computer vision problems. Another interesting problem for the "style detection" problem is the construction of a training set, since, for example, every Dali painting is not "pure Dali", and it will have some areas in it which appear as if they were painted by, say, Van-Gogh. Hence, the training sets of the positive and negative examples respectively will contain some negative and positive elements.

This paper offers a simple, fast, and very easy to implement algorithm which performed reasonably well for the problem of painter identification. The algorithms chooses local features which are based on the DCT transform coefficients, and then classifies the image blocks using the *naive Bayes* classifier, which has proved very useful for text categorization.

2. Previous Work

The scope allotted to submissions must result in injustice when surveying the rich body of work related to image classification. Some references to recent works are given, and these include a more thorough survey. In [1], a mixture model was fitted to the outputs of a filter bank to classify shading and reflectance variations. [8] applies boosting to choose highly selective features for classification. A sophisticated non-parametric multi scale for texture was presented in [2]. An application of global coefficient statistics to noise removal was offered in [7].

3. The Naive Bayes Classifier

The *naive Bayes* classifier is very popular in the data retrieval community, especially in text categorization applications [5, 3]. A short survey of the method's implementation follows.

1. A training set is given, which consists of a set of examples from the categories $\{C_1, C_2, ...C_n\}$. Denote the number of C_i examples as n_i, and the total number as $\sum n_i = n$. The probability of the i-th category is defined as $P(C_i) = n_i/n$. We shall refer to the examples as *texts*, although they do not necessarily have to be textual.

2. Define a set of possible *features*. In textual applications, these are usually words, classes of words which have a similar meaning, or "word stems". A feature may or may not appear in a document. For every feature f_i and category C_j, define $P(f_i/C_j)$ as the ratio

1051-4651/02 $17.00 (c) 2002 IEEE

of C_j's members which contain f_i, and $P(f_j)$ as the ratio of all members in all categories which contain f_j. The important notion of *mutual information* between a feature f_i and category C_j is defined as

$$MI(f_i, C_j) = P(f_i/C_j) \log\left(\frac{P(f_i/C_j)}{P(f_i)}\right) \quad (1)$$

The mutual information has an attractive intuitive meaning; for it to be large, the frequency of f_i in C_j has to large in absolute terms, and it also has to be large relative to f_i's frequency in all the categories.

3. For every category, choose a few features which have the largest mutual information with respect to it. The union of these sets over all categories is called the *feature set*.

4. Given a new text T, extract all the features which it contains – call them $\{f_{i_1}, f_{i_2}...f_{i_k}\}$ – and define for every category C_j the probability that T belongs to it, by

$$P(T/C_j) = \frac{P(T/C_j)P(T)}{P(C_j)} = \frac{P(\{f_{i_1}, f_{i_2}...f_{i_k}\}/C_j)P(\{f_{i_1}, f_{i_2}...f_{i_k}\})}{P(C_j)} \approx \frac{\prod_{l=1}^{k} P(f_{i_l}/C_j) \prod_{l=1}^{k} P(f_{i_l})}{P(C_j)} \quad (2)$$

The first equality is just Bayes' law. The second means that, when classifying T, we only consider the features it contains. The third is an approximation, which assumes that the presence of features is independent (this is where the "naive" in "naive Bayes" comes from); while this is not always true, the technique is still surprisingly effective.

5. Usually, the "non-events" – that is, the non-appearance of a feature in a document – are also considered, which leads to a straightforward extension of Eq. 2.

4. Applying the Naive Bayes Method to Image Classification

The first problem hindering the application of naive Bayes to image classification is: what are the analogues of "text" and "feature" in images? For the task of detecting images which contain some pre-defined structures, one may define a feature as a certain sub-image. For example, for detecting images with human faces, a useful feature would be the presence of an eye in the image. Certain textures can also be recognized by the presence of templates, perhaps up to rotation or scale, etc. Such features, however, are unsuitable for the problem of style detection as presented here (unless we identify a painting by the painter's signature). In general, one cannot hope to base the classification on the presence of a few features in the entire image, because usually painting's cannot be recognized by the presence of a few features in them – the class of features is huge.

Instead, we offer to classify every image block, and classify the entire image by a majority vote. The information extracted from this process contains more than the image classification of the image; it maps the image to different regions, each dominated by a certain style. As will be demonstrated in Section qqq, this often yields results which agree with human intuition. This mapping of the image contains more information than that present in histogram-based approaches, which classify the entire image based on similarity between cumulative distributions of wavelet coefficients.

Also, as opposed to the text categorization applications of naive Bayes, and also to a recent application proposed in computer vision [9], this paper suggests to use *features which have the same size as texts*. We treat each and every image block (the size in the experiments was 9×9) as a text, and the features are the block's 81 DCT coefficients. We say that a certain such feature (coefficient) is contained in a block if its absolute value in the block's expansion is larger than a certain threshold.

5. Implementation and Results

The suggested implementation of the classifier to the problem of painter detection proceeds as follows:

1. Build a database of images. Here, we have tested five painter – Rembrandt, Van-Gogh, Picasso, Magritte, and Dali. Ten paintings from each painter consisted the training set, and the test set consisted of twenty to thirty paintings for each painter.

2. For each DCT basis element (9×9 in size), b_{ij}, and for every artist, the absolute values of the DCT coefficient corresponding to b_{ij} are computed for every 9×9 block in all the artists' paintings in the training set. These values are then binned into 1000 discrete values. This is implemented using a convolution with the respective DCT basis element, hence can be done rather quickly. From this histogram, it is straightforward to construct a table $T(p, i, j, a)$, which stored the probability that, for the painter p, the absolute value of the (i, j) DCT coefficient is greater or equal then a.

3. Naive Bayes requires binary features, so we have to convert the continuous presence of a basis element in a block (that is, its coefficient in the block's expansion),

to a binary one. This is done by thresholding the coefficient absolute value. For every pair of artists and every coefficient, the threshold is chosen so as to maximize the mutual information (Eq. 2). Note that this is a very fast process, once the probabilities of stage 2 have been derived. The maximization is performed over each binned value $\{0, 0.001, 0.002, ...0.999, 1\}$, and over both artists.

4. For each painter, the ten to twenty features with the largest mutual information are chosen. Note that each feature consists of a basis element and a threshold for its coefficient in a block's expansion.

5. Given a new image and a pair of artists, the probability of each image block with respect to each artist is computed from Eq. qqq. We obtained sharper results by considering only blocks with a variance higher than a certain threshold – 20 was a good value, but the results don't change much if 10 or 30 is used. Another heuristic which yielded better results was to classify only blocks for which the winning artists' probability was at least twice than the other artists' probability.

6. Every pixel in the test image is assigned a label, according to the classification of the 9×9 block surrounding it. Pixels whose corresponding window's variance is too small, or for which the ratio between the large and small probabilities does not exceed 2, are labeled as unclassified.

7. The overall classification is determined by a majority vote. However, as noted before, the mapping of individual pixels to different artists contains more information than the overall classification.

5.1. Why DCT?

While we intend to explore other bases – including overcomplete ones – the DCT transform has a property which makes it an attractive candidate for feature selection for the naive Bayes classifier. Recall that the features have to be independent. Since the DCT basis elements are orthogonal, if $(i_1, j_1) \neq (i_2, j_2)$, and b_{i_1,j_1}, b_{i_2,j_2} are the respective basis elements, then under the "natural" probability distribution over images, [4], the random variables $I \to (I, b_{i_1,j_1})$ and $I \to (I, b_{i_2,j_2})$ are independent random variables over the space of images I, which makes them appropriate to use in the naive Bayes paradigm. See discussion of this issue in [2].

5.2. Results

For the five painters tested, a "tournament scheme" classifier was implemented [6]. The rate of success was 86%. Some examples are presented below.

6. Synthesis

Given an image I, we may use the paradigm given here to search for an image which is similar to I but has a high probability that a certain artist painted it. We have used this observation to synthesize paintings ("how would this have looked if Dali painted it?"), but there is not enough space to present the results.

7. Conclusion and Further Research

A simple and very fast algorithm for "image type" classification using the naive Bayes classifier was presented, and applied for the problem of painter classification. Further research will consist of incorporating a multi-level scheme, as well as testing other representations than the DCT. A Markov random field paradigm may be applied in order to create a more consistent segmentation of the image (i.e. not to allow a lonely "Dali pixel" in a "Van-Gogh area" of the image).

References

[1] M. Bell and W. Freeman. Learning local evidence for shading and reflectance. In *ICCV01*, pages I: 670–677, 2001.

[2] J. de Bonet and P. Viola. Texture recognition using a non-parametric multi-scale statistical model. In *CVPR98*, pages 641–647, 1998.

[3] S. T. Dumais, J. Platt, D. Hecherman, and M. Sahami. Inductive learning algorithms and representations for text categorization. In *Proc. 7th International Conference on Information and Knowledge Management CIKM*, pages 148–155, 1998.

[4] D. Keren and M. Werman. Probabilistic analysis of regularization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15:982–995, October 1993.

[5] D. Lewis. Naive bayes at forty: The independence assumption in information retrieval. In *Proc. 10th European Conference on Machine Learning ECML-98*, pages 4–15, 1998.

[6] M. Pontil and A. Verri. Support vector machines for 3D object recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(6):637–646, 1998.

[7] E. Simoncelli and E. Adelson. Noise removal via bayesian wavelet coring. In *IEEE International Conference on Image Processing,*, pages I: 379–382, 1996.

[8] K. Tieu and P. Viola. Boosting image retrieval. In *CVPR00*, pages I:228–235, 2000.

[9] S. Ullman, E. Sali, and M. Vidal-Naquet. Fragment-based approach to object representation and classification. In *Visual Form 2001, 4th International Workshop on Visual Form, IWVF-4, Capri, May 2001*, volume 2059 of *Lecture Notes in Computer Science*, pages 85–102. Springer, 2001.

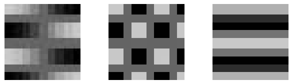

Fig. 1: Three basis elements with largest mutual information for Dali (vs. Magritte). Mutual informations are 0.084,0.079,0.078.

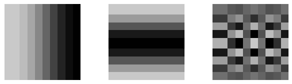

Fig. 2: Three basis elements with largest mutual information for Magritte (vs. Dali). Mutual informations are 0.096,0.075,0.028.

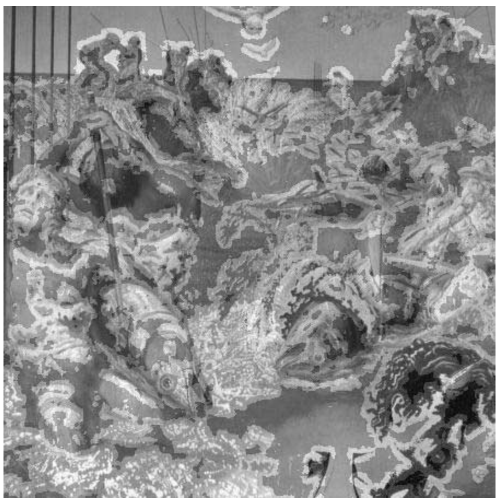

Fig. 3: Excerpt from Dali's "Tuna Fishing". Pixels whose neighborhood was chosen as "Dali" vs. "Magritte" are brighter.

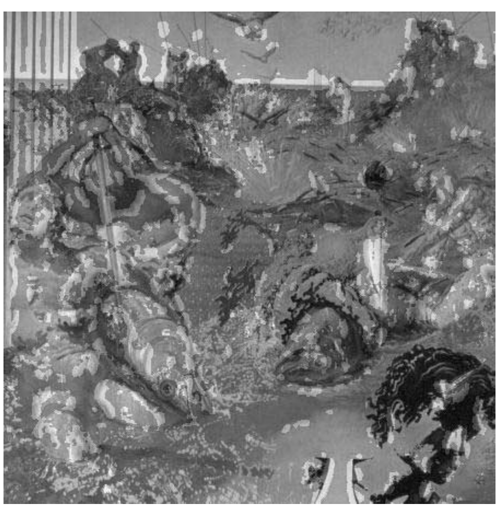

Fig. 4: Excerpt from Dali's "Tuna Fishing". Pixels whose neighborhood was chosen as "Magritte" vs. "Dali" are brighter.

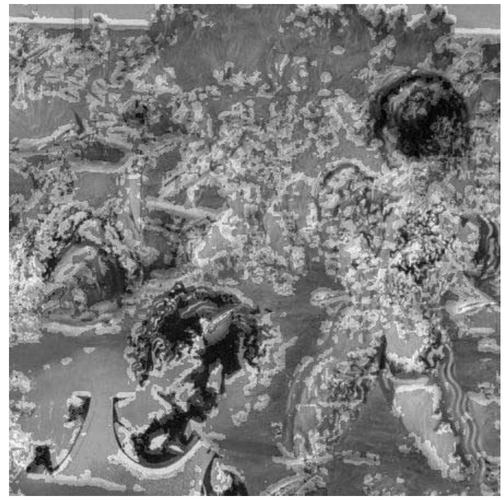

Fig. 5: Excerpt from Dali's "Tuna Fishing". Pixels whose neighborhood was chosen as "Dali" vs. "Van-Gogh" are brighter. Note that the horizon is classified as a "Dali", although it was classified as a "Magritte" in the Dali-Magritte classification. This is consistent with the intuitive notion that "Dali has more straight lines than Van-Gogh, but less than Magritte".

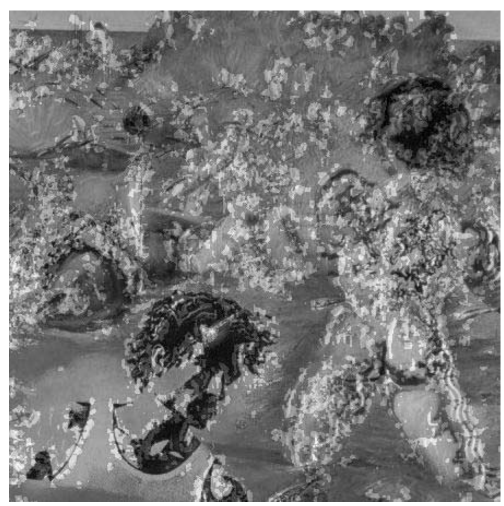

Fig. 6: Excerpt from Dali's "Tuna Fishing". Pixels whose neighborhood was chosen as "Van-Gogh" vs. "Dali" are brighter. Note wavy structures in the right thigh of figure in the right, which are classified as "Van-Gogh", consistently with his "wavy" style.