Who Painted this Painting?

Fahad Shahbaz Khan, Joost van de Weijer and Maria Vanrell

Computer Vision Center Barcelona, Spain

ABSTRACT: Significant research has been made in recent years in the field of automatic object and object category recognition. By describing local regions in images invariant to changes of viewpoint and illumination, computers have shown to be able to reliably retrieve the same object from a data set of images. More recently, people have worked on category recognition, where the task is to recognize an object class (e.g. cars or humans) based on a set of training examples. In this paper, we go one step further and investigate to what extend computers are able to infer the painter from a painting. We will use the bag-of-word approach which basically describes the statistics of small image patches. To test our approach we present a challenging data set of eight different painters. In our experiments color and shape features are used to classify paintings. The results obtained clearly demonstrate the significance of both color and shape features for painting classification.

Fig. 1 Example of painting images. Note that the task is to automatically anotate each painting by its creator (artist).

1 INTRODUCTION: The advent of internet has relinquished a large amount of pictures (here paintings) that are not manageable by humans anymore. Humans have a remarkable ability of recognizing many different object categories in a little time. Although considerable effort has been aimed at making efficient recognition techniques able to classify different object categories robustly, very little work focuses on automatic painting classification problem. Grouping paintings by same artist is a difficult task owing to large amount of variations in style and composition of paintings from the same artist. In order to classify paintings, features related to the pictorial elements of the paintings are required to be extracted. This paper focuses on the difficult problem of visual categorization of paintings by using local features.

There has been a large amount of success in using "bag of visual words" models for object and scene classification [2, 1, 3, 4, 12, 15, 20]. The first stage in the method involves selecting keypoints or regions followed by representation of these keypoints using local descriptors. The descriptors are then vector quantized into a fixed-size codebook. Finally the image is represented by a histogram of the code-book. The image classifier receives histogram representation as an input. The local features play the same role as played by words in traditional document analysis techniques, as they are local and have high discriminative power. A classifier is then trained to recognize the categories based on histogram representations of the images. Due to its success in object recognition, we shall use the "bag of visual words" approach to painting categorisation problem.

Color have been also used previously for many tasks in computer vision especially for object categorization, content-based retrieval, and scene classification etc. Beside color several other features such as texture, shape, and motion have been used to describe visual information for visual categorization. Similary it is further known that combining these local features (i.e. shape and color) provides the best results. In this paper we focus on two types of local features namely color and shape to classify painting categories.

Combining color and shape with in the bag of words approach is still an open research problem. Generally, the fusion of color and shape is carried out in the visual-vocabulary construction stage. Creating a visual vocabulary is a challenging task as the vocabulary should be able to describe widely varying classes. Some classes might have very distinctive color, some very characteristic shape patterns and some might be characterized by combining both features. There exist two main approaches to fuse color and shape into the bag-of-words representation. The first approach, called *earlyfusion*, involves fusing local descriptors together and creating one joint shape-color vocabulary. The second approach, called *latefusion*, concatenates histogram representation of both color and shape, obtained independently. Recently fahad et al. [8] propose a novel way of adding the color information where color is used to guide attention by means of a top-down category-specific attention map employed to modulate the shape words computed using a standard bag-of-words approach. This approach was shown to outperform both early and late fusion. Thus due to its success, we will use Color Attention approach here for combining color and shape features.

It is believed that paintings from the same artist share certain drawing characteristics like same visual patterns or styles. One typical example is that of Picasso's paintings. The style of his paintings is more conventional having an abstract nature to it and include paintings of different styles such as realism, caricature, the Blue Period (where the paintings have blueish hue), and the Rose Period (paintings mostly containing pinks and beiges, light blues, and roses). To this end, in this paper we have introduced a new data set of paintings from eight renowned artists namely, Picasso, Van Gogh, Titian, Ruben,.... There are 40 paintings in each category. The data set is challenging due to large variation in styles of these different artists.

The rest of the paper has been organized as follows. In section 2 some of the related works in this direction is discussed. Afterwards, in section Section 3 our propose approach based on Color Attention framework is presented. Section 4 presents the experimental details like the classification algorithm, the dataset used and the classification settings. Detailed experiments are shown in section 5. Finally, we sum up the conclusions.

2 Related Work: Previously, several works have focussed on the problem of automatic painting classification. Applying different type of brush strokes to classify portrait miniatures has been explored in the work of Melzer et al. [11]. In their approach, both a model based and a semi-parametric, neural network approach has been evaluated and compared experimentally. Sablatnig et al. [13] examined the structural signature based on brush strokes in portrait miniatures. Painting styles of artists are compared by [9] using mixture of 2D multiresolution hidden Markov models (MHMMs) as statistical modeling method for extracting different types of strokes or washes of an artist. The work mainly focus on the problem of Chinese painting classification. Recently Shen et al. [14] propose a framework where multiple features are integrated for the task of automatic classification on large western painting image collections. The work presented in this paper also focuses on the problem of automatic painting classification.

The use of low-level features for painting classification has been investigated in several recent works. Jiang et al. [5] proposed a framework to categorize traditional Chinese paintings. Texture features, color histogram, color coherence vectors, and edge size histogram are used as low-level features. As per learning is concerned, C4.5 decision tree classifier is firstly used as a pre-classification and SVM is used as a final classifier. The work of [7] uses local features based on the DCT transform coefficients followed by the classification of image blocks using the naive Bayes classifier. Similarly in this paper, we also propose a novel framework using local features for the problem of painting classification.

3 Top-Down Color Attention based Image Representation: Human visual system performs effective object recognition using attention mechanisms, which are the strategies to reduce the computational cost of a data-driven visual search . The two ways by which information can be used to direct attention are, bottom-up, where the attention is directed to the salient regions and, top-down, which enables goal directed visual search. Similarly, as in this work we shall use learned color attention to modulate the image representation phase.

Among several properties of visual stimuli, only few are used to control the deployment of visual attention [19]. Color is one such attribute which is undoubtedly used to guide visual attention [19]. Jost et al. [6] measures the contribution of chromatic cue in the model of visual attention. Several other studies performed recently also reveal the importance of color in visual memory [16, 18]. Similarly, in our work color plays a twofold role, firstly, it contains some additional information which is category-specific, and secondly, it modulates the shape words which are computed using a standard bag-of-words approach.

Within the bag-of-words framework each image I_i, i=1,2,...,N contains a number of detected local features f_{ij}, j=1,2,...,M^i . These local features are then represented by the visual words w_i^k, i=1,2,...,V^k and $k \in \{s, c\}$ for the two cues shape and color. The computation of top-down color attention based image representation is done according to:

$$n\left(\mathrm{w}^s|I^i, class\right) = \sum_{j=1}^{M^i} p\left(class|w_{ij}^c\right)\delta\left(w_{ij}^s, \mathrm{w}^s\right) \tag{1}$$

The probabilities $p\left(class|w_{ij}^c\right)$ are computed by using Bayes,

$$p\left(class|\mathrm{w}^c\right) \propto p\left(\mathrm{w}^c|class\right) p\left(class\right) \tag{2}$$

where $p\left(\mathrm{w}^c|class\right)$ is the empirical distribution,

$$p\left(\mathrm{w}^c|class\right) \propto \sum_{I^{class}} \sum_{j=1}^{M^i} \delta\left(w_{ij}^c, \mathrm{w}^c\right), \tag{3}$$

obtained by summing over the indexes to the training images of the category I^{class}. We use the training data for computing the uniform prior over the classes $p\left(class\right)$. By computing $p\left(class|w_{ij}^c\right)$ for all local features in an image, a class-specific color attention map is constructed. The class-specific color attention map is used to modulate the sampling of shape features where in regions with high attention more shape features are sampled than in regions with low attention. This leads to a different distribution over the same shape visual words for each category. The final image representation is obtained by concatenating all the class-specific histograms.

4 Experimental Setup: In this section we discuss the implementation details of the descriptors and detectors used for our experiments and a brief description of the data sets used for the evaluation purpose.

Implementation Details: To test our method for this problem, we use a multiscale grid detector for detecting regions

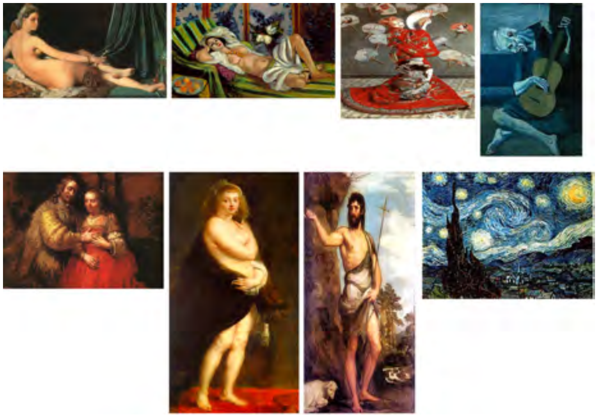

Fig. 2 Typical Examples of Each Class from the dataset. Top row (from left to right): Ingres, Matisse, Monet and Picasso. Bottom row(from left to right): Rembrandt, Rubens, Titian and Vangogh.

from paintings. We normalized all the patches to a standard size and descriptors are computed for all regions in the feature description phase. A visual vocabulary is then computed by clustering the descriptor points using K-means algorithm. In our approach the SIFT descriptor is used to create a shape vocabulary. For color vocabulary, we use Color Name descriptor [17]. Color naming involves the assignment of linguistic color labels to image pixels. The 11 colors names used are black, blue, brown, grey, green, orange, pink, purple, red, white and yellow. Each image is represented by a frequency histogram of visual words. A classifier is then trained based on these histograms. In our experiments we use a standard non-linear SVM with intersection kernel [10]. We compare our method with the individual visual features namely, color and shape.

Image Data Set:

The approach outlined above is tested on a data set with paintings from ten different artists (Ingres, Matisse, Monet, Picasso, Rembrandt, Rubens, Titian, VanGogh). Each category contain 40 images. The dataset has been divided into train set and test set. For training, 25 images from each category are used and the remaining 15 are used for test

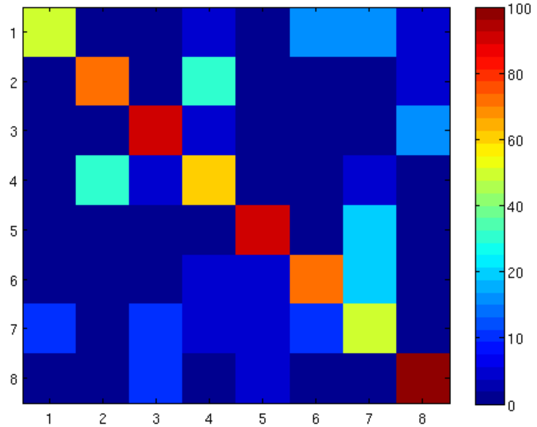

Fig. 3 Confusion Matrix. Note the ambiguity among several categories. For example some of the test images of Matisse are misclassified as Pciasso category.

Fig. 4 Top five images retrived for the first four classes. First row: Ingres, Second row: Matisse, Third row: Monet and fourth row: Picasso. Note that the first 5 images retrived for the the class Matisse conatins images from the category Picasso. The misclassified images are surrounded by a red square.

purpose. The dataset is very challenging due to wide range of painting styles from different artists. Figure 2 shows some of the images from the data set.

5 Experiments: In this section we present the results obtained by using shape features, color features and Color Attention method. Table 1 shows the results obtained from these three methods.

Method	Classification Score
Shape	56
Color	46
Color + Shape	62

Table 1 Classification Score (percentage) using Shape and Color and Combined Color and Shape.

The results in the figure 3 show that shape features perform better than color features. Moreover the Color Attention approach that combines color and shape features outperform the color and shape features alone. This goes to show that in order to achieve best results both color and shape features are required.

Figure 4 shows the confusion matrix for all the eight categories in the data set where the first row/column presents category 1 and so on. The matrix provides information about the misclassified labels (confusion) of different categories. It is noteworthy to mention the ambiguity between categories Matisse and Picasso as both these artists are known to have some similarities in their work. Similary there is a close resemblance between the categories Titian and Ingres. The top five best classified images from the categories Ingres, Matisse, Monet and Picasso are shown in figure 5. As highlighted in the cnofusion matrix, the top five images retrieved for the class Matisse contains images from Picasso category.

6 Discussion and Conclusion: In this paper we present a novel approach for classifying paintings using bag-of-words based approach. We present results based on color and shape features. Moreover we also present results based on Color Attention framework. The outcome of our experiments clearly suggest that combining color with shape using the Color Attention approach provide the best results. As

a future work it would be interesting to observe the results obtained when combining color with other visual features such as texture. Moreover it would be interesting to combine the local feature approach with other approaches that takes into account the composition and style properties of different paintings.

7 Acknowledgement This work is partially supported by the Ramon y Cajal program, Consolider-Ingenio 2010 CSD2007-00018 and TIN2009-14173 of Spanish Ministry of Science, and the Marie Curie Reintegration program of the European Union.

REFERENCES:

[1] A. Bosch, A. Zisserman, and J.Munoz. Scene classification via plsa. In *Proc. ECCV*, 2006.

[2] G. Csurka, C. Bray, C. Dance, and L. Fan. Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision, ECCV*, pages 1–22, Prague, Czech Republic, 2004.

[3] G. Dorko and C. Schmid. Selection of scale-invariant parts for object class recognition. In *Proc. ICCV*, 2003.

[4] L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. In *Proc. CVPR*, 2005.

[5] S. Jiang, Q. Huang, Q. Ye, and W. Gao. An effective method to detect and categorize digitized traditional chinese paintings. *Pattern Recognition Letters*, 27(7):734–746, 2006.

[6] Timothée Jost, Nabil Ouerhani, Roman von Wartburg, René Müri, and Heinz Hügli. Assessing the contribution of color in visual attention. *CVIU*, 100(1–2):107–123, 2005.

[7] Daniel Keren. Painter identification using local features and naive bayes. In *Proc. Int. Conf. on Pattern Recognition*, 2002.

[8] Fahad Shahbaz Khan, Joost van de Weijer, and Maria Vanrell. Top-down color attention for object recognition. In *International Conference on Computer Vision*, 2009.

[9] J. Li and J.Z. Wang. Studying digital imagery of ancient paintings by mixtures of stochastic models. *IEEE Transactions on Image Processing*, 13(3):340–353, 2004.

[10] Subhransu Maji, Alexander C. Berg, and Jitendra Malik. Classification using intersection kernel support vector machines is efficient. In *CVPR*, 2008.

[11] T. Melzer, P. Kammerer, and E. Zolda. Stroke detection of brush strokes in portrait miniatures using a semi-parametric and a model based approach. In *International Conference on Pattern Recognition*, 1998.

[12] P. Quelhas, F. Monay, J. Odobez, D. Gatica-Perez, T. Tuytelaars, and L. Van Gool. Modelling scenes with local descriptors and latent aspects. In *Proc. ICCV*, Beijing, China, 2005.

[13] R. Sablatnig, P. Kammerer, and E. Zolda. Hierarchical classification of paintings using face- and brush stroke models. In *International Conference on Pattern Recognition*, 1998.

[14] Jialie Shen. Stochastic modeling western paintings for effective classification. *Pattern Recognition*, 42(2):293–301, 2009.

[15] J. Sivic, B. Russell, A. Efros, A. Zisserman, and W.Freeman. Discovering object categories in image collections. In *Proc. ICCV*, Beijing, China, 2005.

[16] Ian Spence, Patrick Wong, Maria Rusan, and Naghmeh Rastegar. How color enhances visual memory for natural scenes. *Psychological Science.*, 17:1–6, 2006.

[17] J. van de Weijer, C. Schmid, and J.J. Verbeek. Learning color names from real-world images. In *Proc. CVPR*, Minneapolis, Minnesota, USA, 2007.

[18] Felix A. Wichmann, Lindsay T. Sharpe, and Karl R. Gegenfurtner. The contributions of color to recognition memory for natural scenes. *Journal of Experimental Psychology:Learning, Memory, and Cognition.*, 28:509–520, 2002.

[19] J. M. Wolfe and T.S Horowitz. What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5:1–7, 2004.

[20] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object catergories: An in-depth study. *IJCV*, 73(2):213–238, 2007.