A Recommendation System for Paintings using Bag of Keypoints and Dominant Color Descriptors

Ricardo Ribani and Mauricio Marengoni
PPGEEC, Universidade Presbiteriana Mackenzie, São Paulo, SP, Brazil

Keywords: Image Retrieval, Recommendation, Bag of Keypoints, SURF, Dominant Color Descriptor, Art Paintings.

Abstract: Determining the visual description for a painting is an interesting task that can be used in different applications, like retrieval, classification and recommendation. A painting can differ from others depending on the time period it was painted, the genre and the art movement the author lived. This paper present an approach for content based image retrieval applied to art paintings using the concept of bag of keypoints and SURF detector. A descriptor for dominant color is also used and weighted for a best visual retrieval.

1 INTRODUCTION

With the increasing amount of data and multimedia collections, more and more people find themselves in situations of doubt to make a selection of content, the user accesses a set of information but not all the content is relevant. Recommender systems has received increasing attention for helping the user to find a small amount of information according to your interest (Tkalcic et al., 2010). The first digital recommender systems emerged in the mid 1990s, along with the popularization of the Internet (Adomavicius and Tuzhilin, 2005).

Museums have their assets of paintings available for visitation. Using a recommender system, the visitation of these collections could become more interactive and enjoyable for a regular visitor by showing him paintings of his interest.

This paper presents a technique to retrieve paintings using the concept of bag of keypoints with the SURF algorithm for detection and description of interest points. The collection was divided into three art movements: Classicism, Modern Art and Cubism. Each one with pictures of landscapes and portraits. As the SURF algorithm only evaluates images in grayscale, a dominant color descriptor was used to describe colors. From these two descriptors, it was possible to apply weights to each index, returning images according to the user's interest. The feature points descriptor is based on the characteristics of style and genre, the dominant color descriptor is based on visual similarity by predominant colors.

The paper is organized as follows: in the section 2 related work is presented; in section 3, the art movements and an introduction to art history is presented; in section 4 the methodology, algorithms and tools used are described; in section 5, the results are presented; finally in section 6, conclusion and future work are discussed.

2 RELATED WORK

Several researches in art paintings has focused its efforts in the area of classification and retrieval. Results obtained using the artistic concept of colors for art paintings retrieval shows 79.8 % of accuracy using groups according to the characteristics of colors, without citing movements or genres (Yelizaveta et al., 2005). Another work present a system for classifying art movements where the painting is classified in five movements. A Gabor filter is used for feature extraction in grayscale, a color histogram in HSV space for color descriptor and AdaBoost for machine learning, reaching an accuracy of 68.3 % (Zujovic et al., 2009). A notable result had an accuracy of up to 90 % for classification of art paintings. Six different color descriptors were used together with a Support Vector Machine, classifying art paintings within 3 different art movements (Gunsel et al., 2005).

3 ART MOVEMENTS

Art paintings created in certain art movements such

Ribani R. and Marengoni M..
A Recommendation System for Paintings using Bag of Keypoints and Dominant Color Descriptors.
DOI: 10.5220/0005297603210327
In *Proceedings of the 10th International Conference on Computer Vision Theory and Applications* (VISAPP-2015), pages 321-327
ISBN: 978-989-758-090-1

as Baroque, Realism or Romanticism have similar visual properties. The Baroque movement was developed primarily in Europe between the late 16th century and the mid of the 18th century, the paintings of this movement has a sweeping diagonal element that crosses many planes with sharp contrast between light and dark. The Romantic paintings appeared then in the 19th century and they are very close to Baroque paintings, with realistic elements keeping the diagonal element and the contrast between light and dark to accentuate the dramatic feelings. Still in the 19th century, the Realism movement was created, where artists have learned to use the best scientific knowledge applied in painting and began to leave the emotive vision, seeking to better represent reality. However, the characteristics of the Realism still keeps this movement closed to the Baroque and Romanticism, with soft brush strokes and very realistic appearance (Proença, 2003).

With advances of photography in the mid of 20th century, Modern art was initiated by Impressionist movement which revolutionized painting. In this movement, artists were not seeking to retract perfectly the reality, but spend a certain feeling in his paintings using lighting effects and visible brush strokes. Another feature of this movement is the use of color, compared to previous movements it uses more colors and shades, even due to evolution from paints and color mixing techniques. One of the most important artists of this movement was the french Claude Monet. The movements developed after Impressionism and until the Modern art are called Post-Impressionism and still have the same characteristics. We can cite the Expressionism movement, which began to worry in retract the problems of society but keeping the same artistic elements of Impressionism. Inside of Expressionism we can cite the dutch artist Vincent Van Gogh (Proença, 2003).

Finally we have the Cubism movement, also originated in the 20th century. Cubism seeks to show the objects with all the faces in the same plane and treats the forms of nature through geometric shapes. Among the major artists of this movement, we have Pablo Picasso and Georges Braque (Proença, 2003).

4 METHODOLOGY

To implement the proposed solution, a collection of art paintings was built from www.wikiart.org, a website that contains art paintings made by a public license. The images are organized by artist, genre and movement. Images obtained from the website were analyzed within 6 movements: Baroque, Romanticism, Realism, Impressionism, Expressionism and Cubism. Images in each movements, were divided in landscapes and portraits, forming a collection with a total of 240 images.

To perform a content-based retrieval of the art paintings, the images need to be represented in the form of descriptors, which were extracted by points of interest using the concept of bag of keypoints and a dominant colors descriptor, making possible to pass a query image and index the results according to their similarity.

4.1 Features Extraction

The algorithm used to detect feature points in the images was SURF (Bay et al., 2008), an algorithm based on the same concepts as SIFT (Lowe, 2004). These algorithms are invariant to scale, rotation and partial lighting. Detected feature points will be described and used for generating a visual dictionary using the concept of Bag of Keypoints (Csurka et al., 2004).

SURF uses integral images, which results in a faster processing time when using convolution with box filters. The SURF detector works with a hessian matrix detecting blob-like structures at locations where the determinant is maximum (Bay et al., 2008).

To describe each detected point, SURF creates a vector that describes the intensity distribution in a region neighbor to the considered point, a similar approach on how the gradient information is extracted by the SIFT algorithm. The dominant orientation of the image is extracted from this region, which makes the algorithm invariant to rotation. Each point will be described as a vector of 64 positions, describing how the intensity changes at that point (Bay et al., 2008).

4.2 Bag of Keypoints

Based on the bag of words, the bag of keypoints was presented as a way to quantize local features and classify objects or pictures within a given class (Csurka et al., 2004). The authors addressed the problem of image retrieval in large databases and explained that high level access to information to manage this amount is required, reducing the semantic gap. Based on this principle, the bag of keypoints presents a way to describe and classify each of the images using the local feature points. Detected points need to be clustered to generate a dictionary of visual words. This dictionary will correspond to a histogram with the number of occurrences of a certain pattern in the image (Perronnin, 2008). With an appropriate categorization of the content, it's possible to measure the similarity between images and generate recommendations.

The steps required to build a visual descriptor are:

- Detection and description of the keypoints contained in all images of the dataset using the SURF algorithm;
- Generation of a visual dictionary for each class using the k-means clustering algorithm;
- Count how many times each word appears in the image, resulting in a descriptor vector.

In the original paper describing the bag of keypoints technique (Csurka et al., 2004), the points detected in the first stage are placed in a single group (or bag), then the visual words are generated by the k-means algorithm. But, there is another way to generate the dictionary, the keypoints could be divided into groups and generate one dictionary for each class (Perronnin, 2008).

According to the research done about art history, the art paintings were divided into 3 groups by movements:

- Classicism: Barroque, Romantism and Realism;
- Modern Art: Impressionism and Expressionism;
- Cubism;

Despite the fact that Cubism is part of the Modern art movement, it was placed in another group due the visual characteristics, that are very different from Impressionism or Expressionism.

To generate the visual dictionary, the k-means algorithm was used to receive all detected points and converge then to k centers for each class. Each value indicates a word in the dictionary and is used to generate a histogram that describes visually the image. The value of k must be large enough to distinguish features that classify the image, but not so large to distinguish minor variations such as noise (Csurka et al., 2004).

Having the dictionary defined, the next step consists in identifying how many times each visual word appear in each image (Csurka et al., 2004). The feature points are detected again and each point is assigned to a word in the dictionary using the KNN (K-Nearest Neighbors) technique, each point will represent a word in the dictionary with k nearest neighbors (Marengoni and Stringhini, 2011). A histogram will be created with the same number of positions as the dictionary. For each point the value of the corresponding word is increased, then the values are normalized. This will be the image's describing vector according to the feature points.

4.3 Dominant Color Descriptor

Up to this point we have a descriptor based on the characteristic points generated by the SURF algorithm combined with the bag of keypoints technique. Because the SURF algorithm works only with grayscale images the color characteristic from the paintings were not considered. In order to enhance the image description the color information is required.

The color information is a very important feature in art paintings. The color add beauty to images and provides rich information for content-based image retrieval. A way of index images by color is using the dominant color, which can be represented by fragments of homogeneous color that can be perceived by the human eye (Krishnan et al., 2007). In this work, not only the main dominant color was used, but the eight dominant colors with their percentages.

The generation of the dominant color descriptor consists of three steps:

- Make a color palette based on all images in the collection;
- Quantize colors in each image according to the palette;
- Keep eight dominant colors;

To generate the palette, all images from the collection were processed and the RGB value of each pixel was extracted. These values were included in a single matrix and clustered using the k-means algorithm with a value of $k = 24$. A palette of 24 colors was obtained based on the collection.

With the color palette built, the next step is to quantize the colors of each image. Each pixel in image was assigned to one value of the palette using the KNN technique. An image with a reduced amount of color was obtained, as shown in Figure 2. This image contains a histogram of 24 positions, which shows the percentage of each color.

The 8 biggest values in the histogram were identified and the remaining positions were set to zero obtaining a descriptor with 24 positions representing the 8 dominant colors with respectively percentuals.

Figure 1: Original image with all colors.

Figure 2: Quantized image with 24 colors.

Figure 3: Image with 8 dominant colors.

4.4 Indexing and Retrieval

The task of retrieving images and generating recommendations needs a query image. A method to index images using feature points is to identify and describe all points of all images and store in the database. Then the extraction and description of points in the query image is made. A process of correlation is made between the descriptors of the query and all the descriptors stored in the database, where for each point in the query image there is a point found in the database and the respective image gets a vote. The images are indexed according to the number of votes. Despite the robustness of this model, the correlation uses a brute force algorithm and the processing cost grows proportional to the number of points (Valle and Cord, 2009).

When implementing the bag of keypoints, each image has a single vector of fixed size that describes the distribution of feature points in the image. The color descriptor used in this work, also has a fixed size.

Two different indexes were made, one for the feature points descriptor and other for the color descriptor. For each index a matching was made between the query descriptor and all images in the collection using a brute force algorithm. The L1-Norm or Manhattan distance was used for this, obtaining the distance between each image and query. Then the images were sorted by the distance in ascending order. The query image is always in the collection, so it is expected that the first retrieved image will be the query itself, since the descriptors are equal. The second image retrieved will be the art painting to be recommended to the user, as represented in Figure 4.

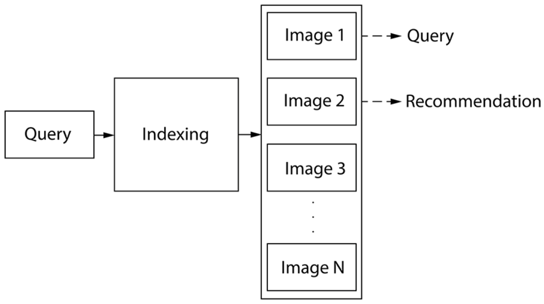

Figure 4: Representation of choice for recommendation from retrieved images.

For considering the feature points and the dominant colors on a common index, the index of each feature were integrated by combining the distance values (Jain and Vailaya, 1996). Let's consider Q as the query image and I an image in the collection, D_p will be the distance between Q and I based on feature points and D_c will be the distance based on dominant colors. The total distance D_t will be:

$$D_t = \frac{w_p D_p + w_c D_c}{w_p + w_c} \tag{1}$$

where w_p and w_c are the weights for feature points and dominant colors, respectively.

5 RESULTS AND DISCUSSION

According to the visual characteristics, the dataset was divided into three groups. Each group has 40 landscapes and 40 portraits as shown in Table 1.

Table 1: Number of imagens included in each class.

Class	Art Movement	Genre	Images
1	Baroque Realism Romanticism	Portrait	40
2	Baroque Realism Romanticism	Landscape	40
3	Expressionism Impressionism	Portrait	40
4	Expressionism Impressionism	Landscape	40
5	Cubism	Portrait	40
6	Cubism	Landscape	40

The first tests were performed to verify the appropriate number of visual words, changing empirically from 250 to 2000 words. For this test the weight of the dominant color descriptor was set to $w_c = 0$. The measurement was made passing all the images in the query, one by one and evaluating the precision for 1 recommendation. As each image has the style and genre information, the accuracy was measured by: style, genre, only one (OR) and both (AND). The results are shown in Table 2.

Table 2: Precision by number of visual words, style and genre.

Words	Style	Genre	AND	OR
250	0,8410	0,6778	0,5941	0,9247
500	0,8828	0,7406	0,6778	0,9456
1000	0,9080	0,8117	0,7448	0,9749
1500	0,9205	0,8117	0,7699	0,9623
2000	0,9247	0,8243	0,7699	0,9791

The best value was obtained using 2000 visual words to at least one feature, but the values are stabilized from 1000 words to any characteristic, as noted in the graphic of the Figure 5. After 1000 words the improvement is not too significant and the computational cost increases with the number of words.

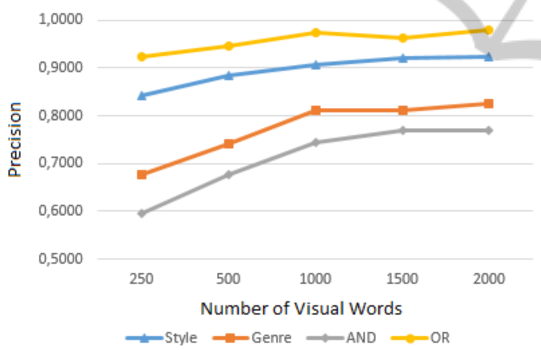

Figure 5: Graphic of precision by number and visual words by style and genre.

When the first recommendation have already been seen by the user, the algorithm will recommend the second image and so on. Thinking about it, the accuracy to recommend a larger amount of images was also measured. 1000 visual words were used for this test and four images were retrieved. The average processing time for the retrieval was 0.038 ms, the results are shown in Figure 6 and the values in Table 3.

Table 3: Precision for 4 recommendations by style and genre.

Style	Genre	AND	OR
0,8588	0,7699	0,6841	0,9446

In the Figure 7 it's possible to see that results are similar according to style or genre but not by color. To evaluate the precision according to the color descriptor, the weight for feature points was set to $w_p = 0$ and the weight for color descriptor was set to $w_c = 1$. Results are presented in Figure 8.

Finally, the weights were changed to integrate the descriptors, this was done empirically. The results were evaluated visually observing the results in relation to the color and taking care to keep good precision values. Thinking about recommendation, these weights can be defined by the user according to his preference, for color or style. The defined values were $w_c = 0.2$ and $w_p = 0.8$. The results for this configuration are shown in Figure 9. Due to the inference of the color descriptor, the accuracy was reduced keeping a precision of 0.6569 for style or genre.

6 CONCLUSION AND FUTURE WORK

The method proposed here presented good results for both recommendation and retrieval of art paintings. Using only the feature points was possible to obtain excellent values of precision. When combining this with the dominant color, we could enhance the visual similarity of the retrieved images. The appropriate division of the paintings was an important step, where the study of the art movements was fundamental. In tests, it was possible to conclude that the size of the vocabulary words is an important choice and should be appropriate according to the type of the images.

It's recommended to perform more accurate performance tests, in order to verify the best dictionary size. Another possible direction for this research is to examine the amount of feature points detected by the SURF in each art painting according to style or genre and make an analysis on these values. To verify if, when reducing that amount of points, it is also possible to reduce the vocabulary size and what the precision rates will be. Finally it's recommended to find a way to describe the spatial distribution of the colors.

ACKNOWLEDGEMENTS

The authors thank Fundo Mackenzie de Pesquisa (Mack-pesquisa) from the Universidade Presbiteriana Mackenzie, CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) for the financial support for this research.

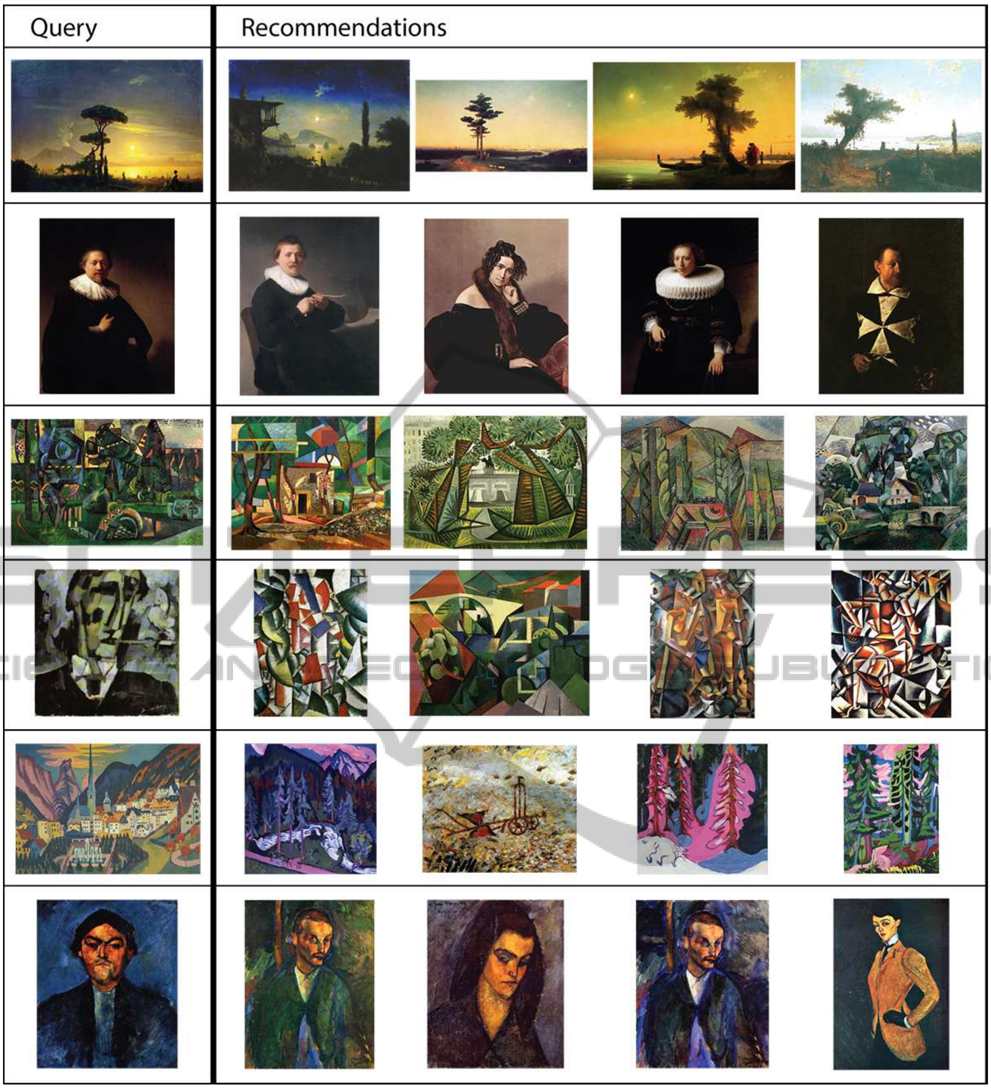

Figure 6: Examples of retrieval results of 4 recommendations ($w_p = 1$ and $w_c = 0$).

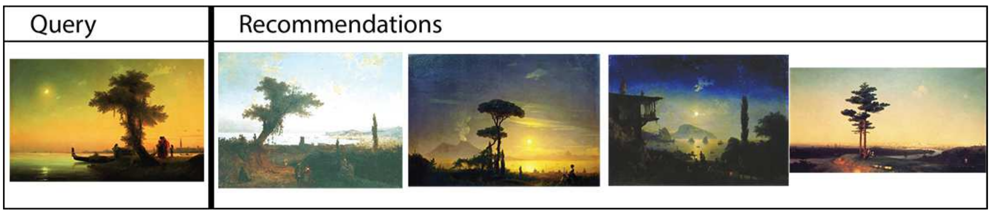

Figure 7: Retrieval results for bag of keypoints descriptor ($w_p = 1$ and $w_c = 0$).

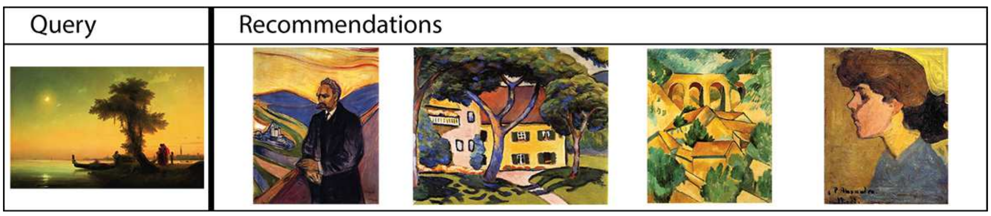

Figure 8: Retrieval results for color descriptor ($w_p = 0$ and $w_c = 1$).

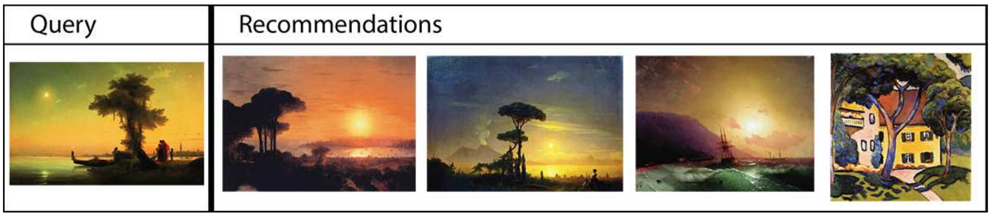

Figure 9: Retrieval results for $w_p = 0.8$ and $w_c = 0.2$.

REFERENCES

Adomavicius, G. and Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. on Knowl. and Data Eng.*, 17(6):734–749.

Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, 110(3):346–359.

Csurka, G., Dance, C. R., Fan, L., Willamowski, J., and Bray, C. (2004). Visual categorization with bags of keypoints. In *In Workshop on Statistical Learning in Computer Vision, ECCV*, pages 1–22.

Datta, R., Joshi, D., Li, J., and Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.*, 40(2):5:1–5:60.

Gunsel, B., Sariel, S., and Icoglu, O. (2005). Content-based access to art paintings. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 2, pages II–558–61.

Jain, A. K. and Vailaya, A. (1996). Image retrieval using color and shape. *Pattern Recognition*, 29:1233–1244.

Krishnan, N., Banu, M., and Callins Christiyana, C. (2007). Content based image retrieval using dominant color identification based on foreground objects. In *Conference on Computational Intelligence and Multimedia Applications, 2007. International Conference on*, volume 3, pages 190–194.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110.

Marengoni, M. and Stringhini, D. (2011). High level computer vision using opencv. In *Graphics, Patterns and Images Tutorials (SIBGRAPI-T), 2011 24th SIBGRAPI Conference on*, pages 11–24.

Perronnin, F. (2008). Universal and adapted vocabularies for generic visual categorization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(7):1243–1256.

Proença, G. (2003). Editora Ática.

Tkalcic, M., Burnik, U., and Kosir, A. (2010). Using affective parameters in a content-based recommender system for images. *User Modeling and User-Adapted Interaction*, 20(4):279–311.

Valle, E. and Cord, M. (2009). Advanced techniques in cbir: Local descriptors, visual dictionaries and bags of features. In *Computer Graphics and Image Processing (SIBGRAPI TUTORIALS), 2009 Tutorials of the XXII Brazilian Symposium on*, pages 72–78.

Yelizaveta, M., Tat-Seng, C., and Irina, A. (2005). Analysis and retrieval of paintings using artistic color concepts. In *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pages 1246–1249.

Zujovic, J., Gandy, L., Friedman, S., Pardo, B., and Pappas, T. N. (2009). Classifying paintings by artistic genre: An analysis of features and classifiers. In *MMSP*, pages 1–5. IEEE.