

Features for Art Painting Classification based on Vector Quantization of MPEG-7 Descriptors

Krassimira Ivanova¹, Peter Stanchev^{1,2}, Evgeniya Velikova³,
Koen Vanhoof⁴, Benoit Depaire⁴, Rajkumar Kannan⁵,
Iliya Mitov¹, and Krassimir Markov¹

¹ Institute of Mathematics and Informatics – Bulgarian Academy of Sciences,
Sofia, Bulgaria

kivanova@math.bas.bg, mitov@mail.bg, kmarkov@math.bas.bg

² Kettering University, Flint, MI, 48504, USA

pstanche@kettering.edu

³ Faculty of Mathematics and Informatics, Sofia University, Sofia, Bulgaria

velikova@fmi.uni-sofia.bg

⁴ IMOB, Hasselt University, Hasselt, Belgium

koen.vanhoof@uhasselt.be, benoit.depaire@uhasselt.be

⁵ Bishop Heber College, India

rajcumarkannan@yahoo.co.in

Abstract. An approach for extracting higher-level visual features for art painting classification based on MPEG-7 descriptors is presented in this paper. The MPEG-7 descriptors give a good presentation of different types of visual features, but are complex structures. This prevents their direct use into standard classification algorithms and thus requires specific processing. Our approach consists of the following steps: (1) the images are tiled into non-overlapping rectangles to capture more detailed information; (2) the tiles of the images are clustered for each MPEG-7 descriptor; (3) vector quantization is used to assign a unique value to each tile, which corresponds to the number of the cluster where the tile belongs to, in order to reduce the dimensionality of the data. Finally, the significance of the attributes and the importance of the underlying MPEG-7 descriptors for class prediction in this domain are analyzed.

Keywords: Content-Based Image Retrieval (CBIR), Multimedia Semantics, MPEG-7 Descriptors, Clustering, Vector Quantization, Categorization.

1 Introduction

The digitalized art painting space gives us the opportunity to immerse in the ocean of accumulated culture. Earlier we could only dream to see some masterpieces. Now our computer moves us to every chosen place and time. These abilities increase the expectations for easy resource discovery by different criteria. While one user could be interested in art paintings from a specific movement or artist, others would search for images with particular theme or composition, and still others can be attracted by some pure aesthetic influence of the image. Many efforts are aimed at combining text-based and content-based search technologies in real-world image retrieval.

MPEG-7 is an ISO/IEC standard developed by Moving Picture Experts Group. The formal name of MPEG-7 is "Multimedia Content Description Interface" (<http://www.chiariglione.org/mpeg>). The standard describes the multimedia content data that supports some degree of interpretation of the information meaning. The elements that MPEG-7 standardizes support a broad range of possible applications. The visual descriptors in the MPEG-7 standard describe different aspects of the image content such as dominant colors, edginess, texture, etc. The MPEG-7 descriptors are often used in the process of image-to-image matching, searching for similarities, sketch queries, etc. [1].

In this paper we apply a vector quantization method on the MPEG-7 descriptors as a possible approach for dimensionality reduction. For obtaining simple visual characteristics we use a clustering procedure over the tiles. The processes of obtaining these attributes are realized in our laboratory for art-painting semantic image retrieval APICAS. The significance analysis of received characteristics for class prediction of the artists' names or movements is conducted in the data mining analysis environment PaGaNe [2].

The rest of the paper is organized in the following way. Section 2 presents our methodology for extracting visual features from MPEG-7 descriptors based on vector quantization. Section 3 describes the software used for our research. Section 4 presents the received results. Finally, some conclusions and future work directions are highlighted.

2 Proposed Approach for Feature Extraction

MPEG-7 descriptors are complex descriptors, which provide a good presentation of different types of visual features. The description of the structure of MPEG-7 descriptors and algorithms is given in [3].

These complex structures need specific processing and cannot be put directly into generic classification algorithms.

In our work we are focused on following MPEG-7 descriptors:

- *Scalable Color (SC)* represents the color histogram in the HSV color space, encoded by a Haar transform. For presenting the image or a selected part, *Scalable Color* needs a vector with 64 attributes;
- *Color Layout (CL)* specifies the spatial distribution of colors using *YCbCr* color space. We use the first quantized *DCT* coefficient of the *Y*, *Cb* and *Cr* components, the next five successive quantized *DCT* coefficients of the *Y* component and the next two quantized *DCT* coefficients of the *Cb* and *Cr* component. Thus, the *Color Layout* vector has 12 attributes;
- *Color Structure (CS)*, which specifies both color content and the structure of the content. The descriptor expresses local color structure in an image by means of a structuring element that is composed of several image samples. We use a vector with 64 attributes for representing the *Color Structure*;
- *Dominant Color (DC)*. We reconfigured the presentation of this descriptor as three vectors, representing distribution of quantized hue, saturation and luminance. Such method is already precisely described and used by us in [4]. After this quantization we receive a vector with 23 attributes (13 for hue + 5 for saturation + 5 for luminance);

- *Edge Histogram (EH)* specifies the spatial distribution of five types of edges in local image regions. *Edge Histogram* descriptor produces a vector with 80 attributes;
- *Homogeneous Texture (HT)* characterizes the region texture using the energy and energy deviation in a set of frequency channels. A vector with 60 attributes is used, presenting *Energy* and *Energy Deviation*.

As a result we obtain a vector, which when taken together results more than 300 attributes. From other side each descriptor needs specific similarity measure.

Some of the MPEG-7 descriptors are alternative. For instance, Scalable Color, Color Layout, Color Structure and Dominant Color concern different aspects of the same phenomenon, i.e. distribution of the color within the image or region. It means that not all descriptors have to be given in the classification process. The first goal of this research is to examine which of them are more convenient to be used for extracting simple visual features for the purposes of class recognition.

Visual attributes can represent global characteristics, concerning whole images, or they can be extracted over the part of the images (specific region or tile of the image). Both approaches have their strengths: global attributes deliver integral temper of the image. Local attributes can capture more detailed information, which characterize the artists' styles or movements' specifics but introduce redundancy for the classifier. For reducing the computational weight and redundancy a possibility is to choose only part of the tiles – only chess ordered tiles, starting from the first tile (1,1) or from the second one (1,2) as well as taking into account only left sided or right sided tiles. In this paper we use all the tiles. The second goal of this research is to examine the trade of between the global and local approach.

In our approach we split the images into $m \times n$ non-overlapping rectangles (tiles). The tiles are marked as (i, j) , where $i \in 1..m$ and $j \in 1..n$. Index i increases from the left tile to the right tile and index j increases from the top tile to the bottom tile of the image.

Let $I = \{I_p \mid p = 1..k\}$ be the observed set of k images. Each image is divided into non-overlapping $m \times n$ tiles and as a result a set of $k \times m \times n$ tiles $T = \{I_p^{ij} \mid p = 1..k, i = 1..m, j = 1..n\}$ is produced.

From the observed set, a subset $LI = \{I_q \mid q = 1..l, l < k, I_q \in I\}$ of representative images is extracted, which is used as learning set. The images of this subset produce a set of $l \times m \times n$ tiles $LT = \{I_q^{ij} \mid q = 1..l, i = 1..m, j = 1..n\}$.

For each MPEG-7 descriptor $X \in \{SC, CL, CS, DC, EH, HT\}$ the algorithm consists of following steps:

- for all tiles in LT , the MPEG-7 descriptor and the corresponding feature vector $X \in \{SC, CL, CS, DC, EH, HT\}$ is calculated;
- the tiles in LT are clustered. We used "repeated bisections". The number of the clusters can be different for different MPEG-7 descriptors, but in this realization for all descriptors an equal number α is used;
- the centroids of each cluster are calculated;

- for each tile a value, which corresponds to the number of the cluster where the tile belongs to, is assigned;
- for tiles which were not in the learning set $IT \setminus LT$, the membership of their centroids is calculated using L_1 -metric, and the number of the corresponding cluster is assigned as a value of the tile.

As a result, each image of I is represented with a feature vector with $x \times m \times n$ numerical attributes, where x is the number of MPEG-7 descriptors. For instance in case of using all MPEG-7 descriptors for 3×3 tiling, the number of attributes in this vector is $6 \times 3 \times 3 = 36$. In case of selecting only a subset of the available tiles (chess order or left/right side), the number of features reduces.

A specific of this approach is that received attributes are nominal. The main purpose of the prepared datasets after implementing this approach is to examine the significance of the attributes and the local/global trade-off for class prediction.

3 Realization

In order to realize the proposed algorithm, our laboratory for art-painting semantic image retrieval "Art Painting Image Color Aesthetic and Semantic" (APICAS), firstly presented in [5], was extended with additional functions.

For obtaining the MPEG-7 descriptors our system refers to Multimedia Content Management System MILOS [6].

As clustering algorithm the program *vcluster*, which is part of the CLUTO open source software package [7], is implemented in the system. As clustering method we have used the default method "repeated bisections". The similarity between objects is computed using the correlation coefficient. As clustering criterion function, cluster maximization of sum of square root of sums of similarities of vectors, belonging to given cluster, is used. In this method, each cluster is represented by its centroid vector and the goal is to find the clustering solution that maximizes the similarity between each vector and the centroid of the cluster to which it is assigned.

As knowledge analysis and testing environment, we used the data mining analysis environment PaGaNe [2]. We have used PGN-1 classifier and implemented statistical analyzing tools for checking up our results and extracting regularities for artists' and movements' styles based on the extracted attributes.

4 Experiments

For our experiments we have used datasets that include 600 paintings of 18 artists from different movements of West-European fine arts and one group, which represent Orthodox Iconographic Style from Eastern Medieval Culture (Table 1). The pictures were obtained from different web-museum sources using ArtCyclopedia as a gate to the museum-quality fine art on the Internet as well as from different Eastern public virtual art galleries and museums for extracting Icons.

Table 1. List of the artists, which paintings were used in experiments, grouped by movements

Movement	Artist
Icons (60)	Icons (60)
Renaissance (90)	Botticelli (30); Michelangelo (30); Raphael (30)
Baroque (90)	Caravaggio (30); Rembrandt (30); Rubens (30)
Romanticism (90)	Friedrich (30); Goya (30); Turner (30)
Impressionism (90)	Monet (30); Pissarro (30); Sisley (30)
Cubism (90)	Braque (30); Gris (30); Leger (30)
Modern Art (90)	Klimt (30); Miro (30); Mucha (30)

We have made several experiments with different parameters for:

- the number of tiles – we have ran experiments with equal tiling over width and height from 1×1 (the whole images) up to 7×7 (i.e. $m = n \in \{1, \dots, 7\}$);
- the number of clusters – the experiments were carried out with 20, 40 and 60 clusters;
- as class value, we have used "artists' names" and "movements".

As the set of attributes we have used $\{X(i, j)\}$, where $i = 1 \dots m$, $j = 1 \dots n$, X is SC, CL, CS, DC, EH, HT .

We have processed the datasets under the procedures of attribute selection in order to receive the order of significance of attributes for prediction. We have implemented several attribute evaluation methods – Chi square; Filter; Information Gain, and they give almost equal results for ordering the significance of the attributes, diverging the order for some attributes only in 2-3 positions. Later analyses are made using Chi square evaluation method.

We have summarized the obtained order of attributes by types of descriptors and positions of the tiles by width, respectively by height.

Fig. 1 shows the distribution of significance of MPEG-7 descriptors for class prediction. As it is shown, the *Color Structure (CS)* descriptor is the most informative for our datasets. The dominance of the features, based on the color descriptors (CS, SC, DC, CL) is obvious, which leads to these assumptions:

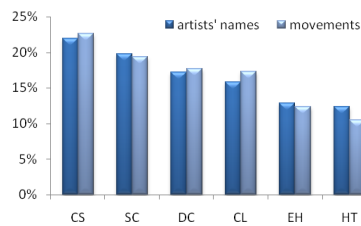


Fig. 1. Average distribution of significance by MPEG-7 descriptors over datasets, using different tiling and clustering, using Chi square evaluation method

- the artist's palettes, which are captured in color descriptors, are a powerful tool for creating the profiles of art painting images;
- using this approach, texture descriptors (EH, HT) can not produce sufficient quality attributes to present the specifics of the brushwork of the artists.

The analysis of the significance of the tile position for the horizontal splitting (Fig. 2) and vertical splitting (Fig. 3), shows that the increasing the number of tiles reduces the differences between the zones.

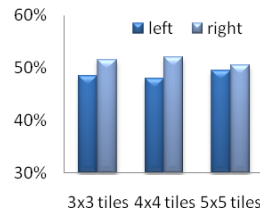


Fig. 2. Distribution of significance of left side and right side of the images for 3×3, 4×4 and 5×5 tiling

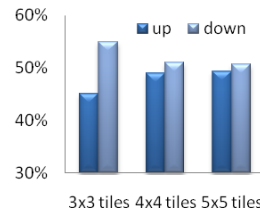


Fig. 3. Distribution of significance of upper and lower zone of the images for 3×3, 4×4 and 5×5 tiling

Furthermore, we can examine the significance of the tiles in respect to the position in the images. Fig. 4 shows the distribution of significance of the tiles in respect to the horizontal position. As we can see, the differences are not distinctive, which means that horizontally each zone of the image is relatively equal important.

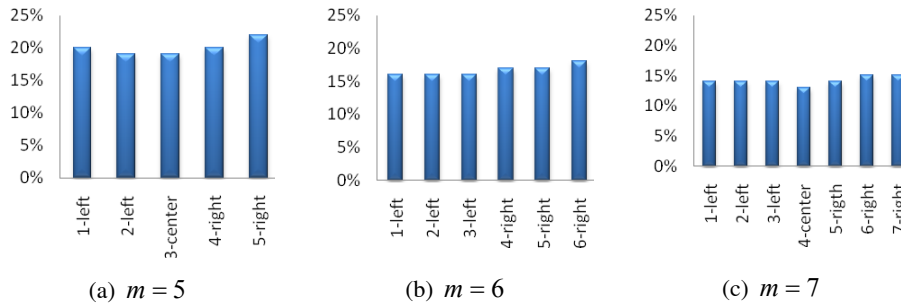


Fig. 4. Distribution of significance of the tiles by position of *width*, $i \in 1 \dots m$ from left to right

The construction of many classical paintings is based on central symmetry. A little superiority of the right part of the image for a fewer number of tiling confirms the results from psychological theories for understanding human perception [8]. We attend to use this fact in further investigation with analyzing the tiles only of the right half of the image.

Similar analysis in respect to the vertical position of the tiles is shown on Fig. 5. Here, it becomes clear that outer tiles (and especially border tiles) are more informative (more distinctive for different classes) than inner tiles (and especially center tiles).

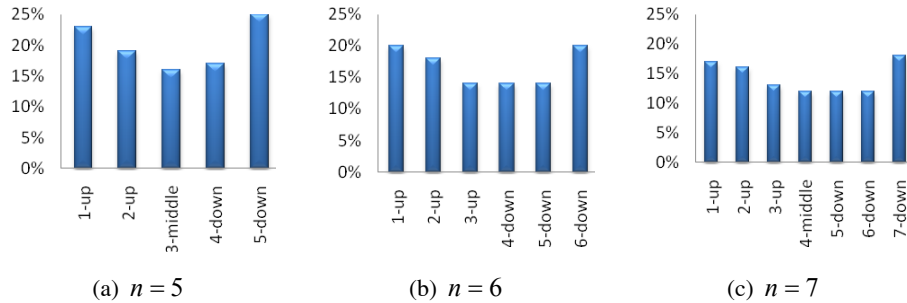


Fig. 5. Distribution of significance of the tiles by position of *height*, $j \in 1 \dots n$ from up to down

This fact can also be explained with differences between constructions of composition in different styles [8]. While the central part of the image brings object or scene information, the borders are less burdened with this task. In order to supply the focus of the image, there are not usually specific objects found here, but only the ground patterns, which are specific for the artists or the school, in which the artists belongs. These patterns capture the ground of the artists' palette and brushwork.

Other experiments are focused on establishing the appropriate number of clusters in order to receive good classification results with lower computational cost. We have run ten-fold cross-validation over the datasets.

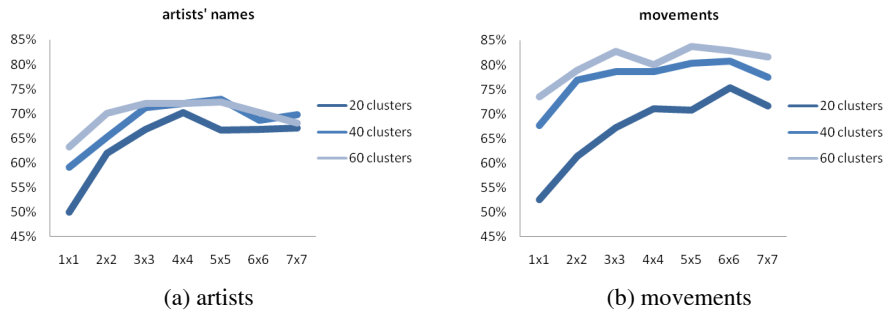


Fig. 6. Classification accuracy for datasets with classes artists/movements using 20, 40, 60 clusters and different numbers of tiling

The results displayed on Fig. 6, show that using vector quantization on MPEG-7 descriptors for the entire image (i.e. 1×1 tiling) is not so informative. Better but non-sufficient results are obtained for 2×2 tiling. Tiling 3×3 is the first with relatively good results. This result is also conceptually validated by the fact that 3×3 tiling corresponds to a rough approximation of the golden ratio, which usually lies in the compositions of the art paintings. The last observed tiling 7×7 shows a decreasing of accuracy, which can be explained by the fact that the pictures became too fragmented and the clusters fall not in the proper positions.

5 Conclusions

In this article we have presented an approach for analyzing visual characteristics, based on MPEG-7 descriptors. The method allows on one hand tiling and on the other hand a huge reduction of the dimensionality of the data. For the used data sets, 4×4 tiling and 40 clusters seemed optimal. From the analysis we learned that the artist's palettes, which are captured in color descriptors, are a powerful tool for creating the profiles of art painting images. The texture descriptors (*EH*, *HT*), in our particular presentation, can not produce sufficient quality attributes to present the specifics of the brushwork of the artists. Position analysis confirmed other research results.

In further research, we will look for combinations of characteristics in order to form profiles for artists or movements.

Acknowledgments. This work was supported in part by Hasselt University under the Project R-1875 "Search in Art Image Collections Based on Color Semantics" and by the Bulgarian National Science Fund under the Project D002-308/19.12.2008 "Automated Metadata Generating for e-Documents Specifications and Standards".

References

1. Stanchev, P., Green Jr., D., Dimitrov, B.: Some Issues in the Art Image Database Systems. *Journal of Digital Information Management*, Volume 4, Issue 4, December 2006, pp. 227-232 (2006)
2. Mitov, I., Ivanova, K., Markov, K., Velychko, V., Vanhoof, K., Stanchev, P.: PaGaNe – a Classification Machine Learning System Based on the Multidimensional Numbered Information Spaces. *Fourth Int. Conf. Intelligent Systems and Knowledge Engineering (ISKE)*, 27-28.11.2009, Hasselt, Belgium, Printed in *World Scientific Proceedings Series on Computer Engineering and Information Science*, No:2, pp. 279-286 (2009)
3. International Standard ISO/IEC 15938-3 Multimedia Content Description Interface – Part 3: Visual, http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=34230
4. Ivanova, K., Stanchev, P.: Color Harmonies and Contrasts Search in Art Image Collections. *First International Conference on Advances in Multimedia (MMEDIA)*, 20-25.07.2009, Colmar, France, pp. 180-187 (2009)
5. Ivanova, K., Stanchev, P., Dimitrov, B.: Analysis of the Distributions of Color Characteristics in Art Painting Images. *Serdica Journal of Computing*, Volume 2, Number 2, Sofia, pp. 111-136 (2008)
6. Amato, G., Gennaro, C., Rabitti, F., Savino, P.: Milos: a Multimedia Content Management System for Digital Library Applications. *Research and Advanced Technology for Digital Libraries: Volume 3232/2004 in Lecture Notes in Computer Science*, Springer Berlin/Heidelberg, pp. 14-25 (2004)
7. Karypis, G.: CLUTO: A Clustering Toolkit Release 2.1.1. University of Minnesota, Department of Computer Science, Minneapolis, MN 55455, Technical Report: #02-017 (2003)
8. Arnheim, R.: *Art and Visual Perception: A Psychology of the Creative Eye*. University of California Press, Berkeley (1974)