Segmentation and classification of fine art paintings

Zuzana Haladova *

Supervised by: Elena Sikudova[†]

Faculty of Mathematics, Physics and Informatics
Comenius University
Bratislava / Slovakia

Abstract

Since the development of the first text-based image search on the internet, the area of image retrieval has come a long way to sophisticated content based image retrieval systems. On the other hand, the semantic gap causes that it is still not possible to create a system which can correctly identify any object in the image. However, this paper proposes a solution for classifying the one sort of objects - paintings. This approach includes segmentation of the painting from the image, creation of the descriptor file from the segmented painting, and classification of the painting by matching its descriptor file to the created database of descriptor files of original paintings. The segmentation of the painting is achieved with 3 preprocessing steps, followed by adjusted Hough transformation. For the estimation of key points and creation of the descriptor file, the SIFT (Scalable Invariant Feature Transform) or the SURF(Speeded Up Robust Features) technique is used. The performance of both techniques is validated within the paper. The solution proposed in this paper was tested on the database of 100 Rembrandt Harmenszoon van Rijn's paintings.

Keywords: Fine art paintings, Segmentation, Classification, SIFT, SURF

1 Introduction

This paper interlopes three scientific areas, image retrieval and the classification and the digital preservation of art. The approach proposed in this paper consists of a CBIR (Content based image retrieval system) which operates over a database of fine art paintings. This CBIR system is defined (according to the categorization proposed by Data et al. [6]) as a system operating over domain specific collection, which queries by an image, with content-based processing of the query.

The motivation for the creation of the system are deficient retrieval possibilities in the most art web-galleries. In these one can retrieve the painting only by its name, which causes problem when you want to find the beautiful painting which you have photographed in the gallery (and immediately forgot the name). This system requires a photograph of a painting on the input and returns the name and the author of the painting found on the photograph. The system works in 2 phases. Firstly, it segments the region consisting of the painting and the frame from the photograph. Segmentation is done by different methods which includes different preprocessing steps, edge enhancement methods followed by the Hough transform [14] or watershed transformation. In the next step the corresponding painting from the database of originals is retrieved. The retrieval is done by comparing the descriptor file of the segmented region created using SIFT [10] or SURF [2] algorithms with the descriptor files of the paintings stored in the database.

This paper is organized in the following way: In the first section the works of other authors in the area of the classification and the digital preservation of art are presented. In the second section the datasets used in this paper for testing and verification are presented. In the third section the process of the segmentation is detailed. In the fourth section the classification methods are recounted. In the fifth section the paper presents the comparison of different methods and the sixth section concludes the paper.

2 Previous work

Since the 80's the computer graphics and vision community is focusing on the problem of the preservation of the cultural heritage. This big mission includes the restoration and the classification of the fine art painting. In this area the most significant assignments are the digital restoration of the paintings, classification of the author's style and categorization of paintings based on the style [8], distinguishing paintings from real scene photographs [5] and determination of new features for paintings classification (Example: Description of painting's texture using brush strokes [15]). For the relatively complete overview see Lombardi [9].

Works in the area of the classification of paintings are mostly focused on author's style or iconography. However one paper shares the assignment with this paper. In the paper [4]a group of students from Stanford were clas-

---

*zhaladova@gmail.com
†sikudova@sccg.sk

sifying the paintings from photographs they took in the Cantor Arts Center. Photographs were taken under constant lighting, without any distractive elements covering the painting. Under these conditions segmentation of the painting from the photograph could be achieved with simple thresholding method. For the classification they used the matching of the color histograms of the photographs with the database of the color histograms of originals.

3 Datasets

For testing and verification of segmentation and classification methods two different datasets were used. Both datasets consist of images of the paintings created by Rembrandt Harmenszoon van Rijn (Figure 1).

The first dataset, (the Originals), consists of 15 photographs of paintings obtained from the Olga's gallery [12], the internet gallery with over 10.000 works of art. These photographs contain the paintings without a frame or a wall photographs are in the resolution 600 times 600 dpi.

The second dataset, (the Photographs) includes 100 photographs taken in museums or galleries by tourists with unspecified digital cameras. This dataset contains photographs from the collection of the author of this paper, from the initiative on her website [1] and from the travel.webshots [1] web portal. Photographs are in different resolutions, miscellaneous scales and are taken under varying lighting. In 8 images the painting is partly covered by the bodies of tourists.

All of the photographs from both datasets are resized to the width of 600px and converted to gray scale. Conversion to gray scale was done by eliminating the hue and saturation information while retaining the luminance from the HSV representation of RGB values of the image.

4 Segmentation

The goal of the segmentation phase was the segmentation of the painting and its frame in the input image (from the Photographs dataset). Three different techniques were used. The primal one used the Gauss gradient method [7], in the improved method the Anisotropic diffusion [13] was applied and the additional method is based on the watershed transformation [11]. Results of the three different methods of the segmentation are presented in the Conclusion section.

---

[1]http://members.chello.sk/halada-j/diplomovka.html

Figure 1: Sample images from the Originals dataset (left) and the Photographs dataset (right). Photographs of the Jewish bride painting in the first row, and of the painting: Portrait of a Young Man in the second row.

4.1 Gauss gradient method

In the primal method the image is processed using Gauss gradient function which computes the gradient using first order derivative of the Gaussian. It outputs the gradient images Gx and Gy of the input image using convolution with a 2-D Gaussian kernel. In the next phase the Gx and Gy gradient images are send as an input to the Hough transform. The Matlab's implementation of the Hough transform is used, since it enables to count the lines from the Hough peaks directly and to connect or trim them based on their length. Lines created in the previous step are then expanded to the borders of the image and lines with big slope are filtered. Lines are then divided into four groups, one for upper, lower, left and right edges. Consecutively, the painting is segmented as the smallest quadrilateral created from the lines. Gauss gradient method is depicted in the figure 2.

4.2 Anisotropic diffusion method

In this approach firstly the histogram equalization is done. Then the image is processed using Anisotropic diffusion, the technique which smooths the image, but preserves the edges. The function is used with the following parameters: number of iterations = 10, kappa = 30, lambda= 0.25 and option = 1 (kappa controls conduction as a function of gradient, lambda controls speed of diffusion, it is 0.25 for maximal stability, option = 1 means the Diffusion equation, this choice favors high contrast edges over low contrast ones). The output image of the function was then convolved with the horizontal and vertical Sobel edge filter, resulting in two binary images Sx, Sy. The images

(a) The input image   (b) Gx image   (c) Gy image

(d) Lines created using Hough trans-
form   (e) Final segmentation
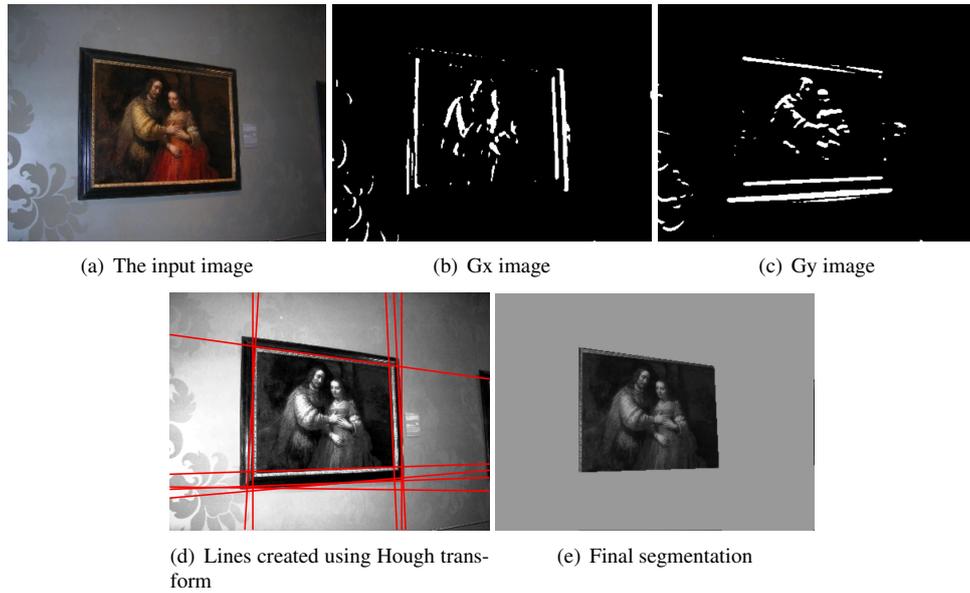
Figure 2: The process of the Gauss gradient method of the segmentation.

(a) The input image I,   (b) Top hat   (c) Bottom hat   (d) (I+ tophat)-bottomhat

(e) Extended minima of (d)   (f) Minima imposition from the com- (g) Clusters created with watershed   (h) Final segmentation
plement of (d) with the marker (e)   transform
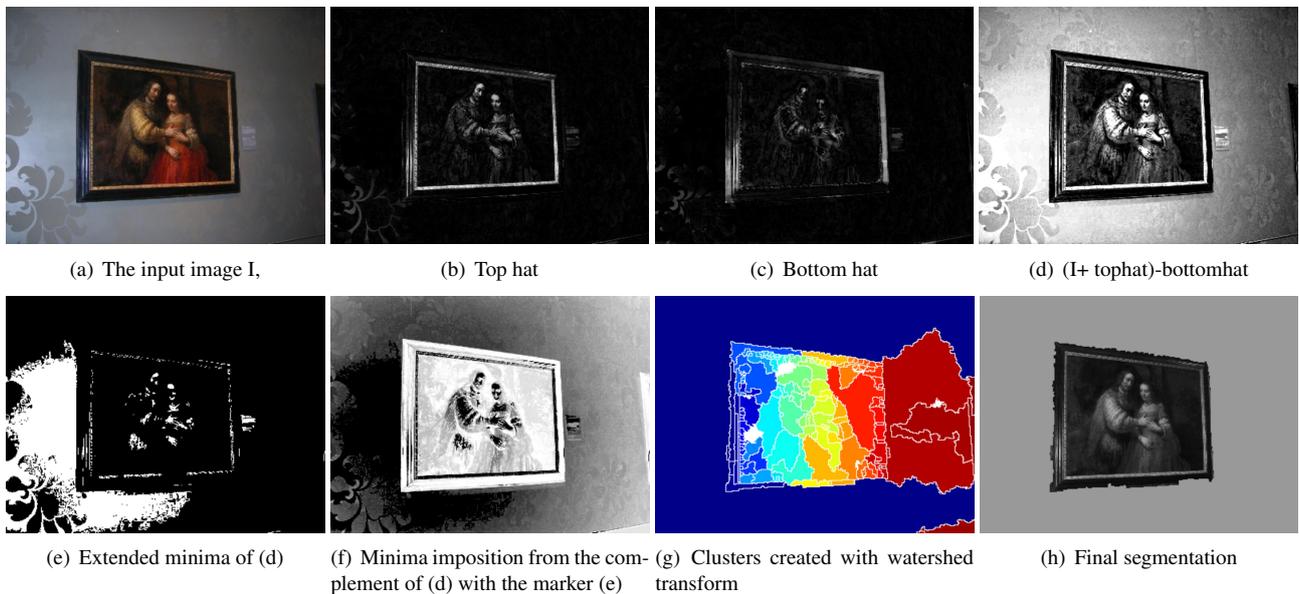
Figure 3: The process of the watershed method of the segmentation.

(a) The input image

(b) The input image processed with Anisotropic diffusion

(c) Sx image

(d) Sy image

(e) Lines created using Hough transform

(f) Final segmentation
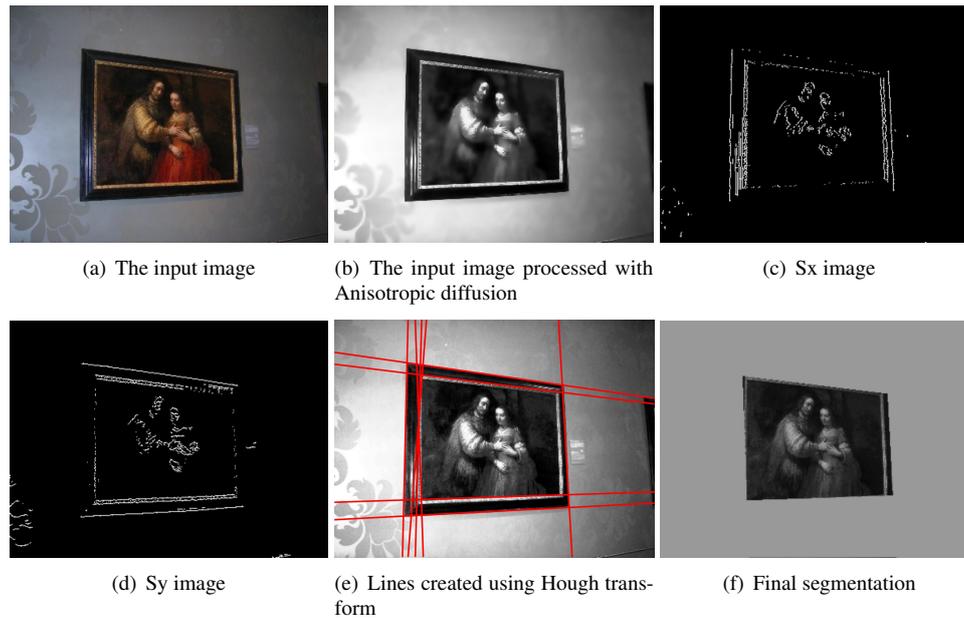
Figure 4: The process of the Anisotropic diffusion method of the segmentation.

Sx, Sy are processed equally to Gx and Gy images in the first method, and also the input image is segmented in the same way. Anisotropic diffusion method is presented in the figure 4.

4.3 Watershed method

In the third approach the input image is firstly preprocessed to enhance edges of the painting's frame. Afterwards the watershed transform is applied. The preprocessing phase consists of 4 steps: 1. Create top and bottom hat of the input image. 2. Create image I2= (I+ tophat)-bottomhat. 3. Create Im3 as extended minima (regional minima of the H-minima transform) of Im2. 4. Im4 is created as the minima imposition from the complement of Im2 with the marker Im3. In the next step, clusters are created with watershed transform applied on the Im4 image. In the last phase the final segmentation is made by growing the background from the corners in the clustered image. Watershed method is presented in the figure 3.

5 Classification

The process of the classification of the painting segmented from the input image is divided in two steps. First the database of descriptor files of the Originals is created. Then the descriptor file of the painting is matched with the database to find the corresponding original. The descriptor file is a N by M matrix, where N is number of the interest points found in the image and M is the length of the descriptor (128 values for SIFT and 64 for SURF). The two different methods are used for producing the de-

scriptor files. SIFT (Scalable Invariant Feature Transform) developed by D. Lowe [10] and SURF (Speeded Up Robust Features) developed by H. Bay et al. [2].

5.1 SIFT

Sift method consists of a detector and a descriptor of features invariant to translation, rotation, scale, and other imaging parameters.

In the first step of the method the interest points (IPs) are identified in the image by the detector. Then the descriptor for each IP is created. The detector is localising the IPs in the scale-space pyramid, which is created by the consequent scaling of the image, its filtering with the Gaussian kernel and the substraction of subsequent filtered images in each scale. IPs are chosen as the local extremes in the 3x3x3 neighbourhood.

The descriptor for each IP is summarized from the orientation histograms of 4x4 subregions of the IP neighbourhood. In every sample point of the subregion the size and the orientation of the gradient is computed and weigted by the Gaussian window indicated by the overlaid circle. Orientation histogram with 8 directions is created from these values. The descriptor of IP than consists of 8 values for all 16 subregions (128 values).

5.2 SURF

SURF likewise SIFT includes the detector and the descriptor. SURF also operates in the scale-space for identifing the IPs, but unlike SIFT it convolve the original image with the different scales of the box filters (approximations of the Gaussian second order partial derivatives in the y a
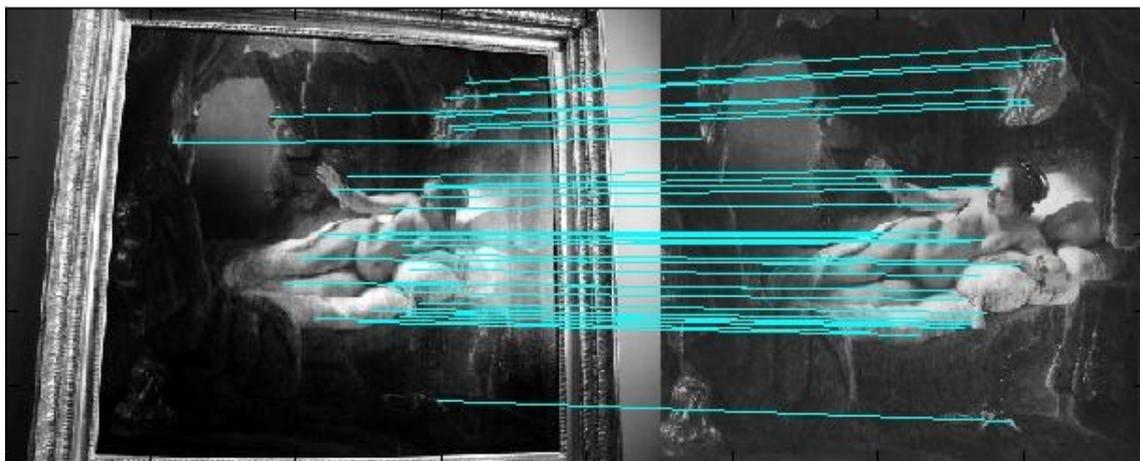
Figure 5: Correspondence of interest points between two paintings matched with SIFT. Painting Danae from the Photographs dataset on the left side and Danae from the Originals on the right side.

xy directions). In order to localise IPs in the scalespaces, a non maximum suppression in a 3x3x3 neighborhood is applied.

64 valued descriptor is created in a few steps. Firstly, the dominant orientation of the IP is extracted from the circular neighborhood as the longest vector estimated by the calculating the sum of all response (the Haar-wavelet responses in the x and y direction, weighted by the Gaussian window) within a sliding orientation window covering the angle of $\frac{\Pi}{3}$. Than, the square region around the IP is created and oriented along the dominant orientation. Lastly, the region is divided into 4x4 subregions. In every subregions 4 features are counted from 5x5 uniformly distributed points. These 4 features are $\sum dx$, $\sum dy$, $\sum |dx|$, $\sum |dy|$: sum of Haar-wavelet responses in the horizontal and vertical direction and the sum of absolute values of Haar-wavelet responses in the horizontal and vertical direction. Four features for all of 16 regions produce the 64 values for every IP.

5.3 Matching

In the matching phase of the classification the binary representations of the descriptor files from the database are loaded and matched with the descriptor file of the segmented painting (DF1) using the nearest neighbor technique. In both SIFT and SURF approaches for each descriptor file (DF2) from the database, the value of the matching with DF1 is counted. For every row (corresponding to the descriptor of one IP) of DF1 the nearest neighbor and the second nearest neighbor from DF2 is counted. Nearest neighbor is a row from DF2 with the smallest Euclidean distance from the DF1 row. The matching value is then the sum of DF1 rows for which the nearest neighbor has value smaller than distanceRatio times second nearest neighbor. The distanceRatio was determined by authors of SIFT and SURF as 0.6 and 0.7. The painting from the Originals dataset with the greatest matching value is

elected as the painting best corresponding to the input image.

It is possible that two non corresponding paintings have the matching value greater a 0. This is caused by the fact that features recognized by the SIFT and SURF are not distinctive at 100% and additional inaccuracies are caused by the blur and the noise. In order to prevent incorrect classification of the paintings not present in the Originals database, the threshold for minimal matching value is established. If the DF2 best corresponding to the DF1 of the input painting has the matching value smaller than the treshold, the input painting is not present in the Originals dataset.

5.4 Other methods

In the scope of this paper, one additional method for creating descriptors was verified - the colorSIFT developed by G. J. Burghouts and J. M. Geusebroek [3]. The color extension to the original SIFT considers color gradients, rather than intensity gradients, in the Gaussian derivative framework. This approach, however, proved to be ineffective for our purpose. In the 10 tested images, there was found only 10% of the matches found by SIFT.

6 Results

This section summarizes the results of each step of the proposed system.

6.1 Segmentation results

In the segmentation phase the methods were tested on the Photographs dataset (Images from the Originals dataset are already segmented). This dataset consists of the

Method	Gauss gradient	Anisotr. Diffusion	Watershed
Correct segmentation	73%	89%	49%
Over segmentation	6%	3%	1%
Under segmentation	21%	8%	50%

Table 1: Percentage of paintings properly segmented by different methods

photographs taken by tourists in different galleries, under different lighting condition and with different cameras. Within the segmentation phase most problems were caused by the low contrast of the photographs, which was eliminated in the Anisotropic diffusion method by the equalization of the histogram. Table 1 summarizes the percentage of the correct segmented versus over and under segmented paintings.

Over segmentation, mostly in the Gauss gradient method was induced by the strong edge responds in the paintings, especially in the painting Night Watch (Rijksmuseum, Amsterdam) where the pale flags and spears has very strong color edges in the black background. Other problem with the Night Watch was the low contrast of the black frame of the painting to the dark gray wall paint. In the Watershed method, over segmentation occurs in one image, where the shadow in the upper right side of the image blend with the black upper right corner of the painting. Under segmentation arises, when the paintings frame is mostly covered or in low contrast with the wall or the background of the painting contains strong edges (wall corner or cartouch presented on the photograph).

The problems with over and under segmentation were partly eliminated by using the Anisotropic diffusion in the second method, which smoothes the color edges in the painting and also the edges in the background, but preserve the edges of the frame. The primal method, the Gauss gradient uses smoothing with the Gaussian kernel, which smoothes all edges uniformly. The Watershed method was integrated to present a different approach to the segmentation, but the results indicates that it is not efficient for this purpose. Finally, as expected the best results were achieved with the Anisotropic diffusion method (See table 1).

6.2 Classification results

Two methods were used for the classification purpose, SIFT and SURF. In this section both methods will be evaluated in a sense of a precision and the speed. For the evaluation purpose both datasets were used. One hundred images from the Photographs dataset were divided into 16

Method	SIFT	SURF
threshold = 0	75%	73%
threshold = 6	88%	90%
threshold = 8	89%	88%
threshold = 12	90%	82%

Table 2: Percentage of properly classified paintings by SIFT and SURF methods with different thresholds

Method	SIFT	SURF
time of the computation of one descriptor file	0,8125 s	0,32025 s

Table 3: Time spent on the computation of one descriptor file with different methods

groups, 15 groups corresponding to 15 originals present in the Originals dataset and one group with the images of the paintings not presented in the Originals dataset. The descriptor file from segmented paintings was created and the best matching original was chosen. Painting was labeled with the number of the best matching original and the label was compared with the number of the image's group. If the best matching original has matching value smaller then a threshold (the threshold was established on the value 7 for SIFT and 6 for SURF), the image was labeled with the 16- not presented in the database. Correct classification means that image was labeled with the number equal to the number of its group. Table 2 present the number of correctly classified images with the SIFT and the SURF methods. Additional value for the performance measure is the time of the computation of one descriptor file in Matlab. The time value is presented in the table 3.

7 Conclusion

As a conclusion the best results in the segmentation and the classification of fine art paintings from photographs were achieved with the combination of Anisotropic Diffusion and SURF methods. Both SURF and SIFT achieved 90 % percent of succesfully classified paintings, which means that 90 photographs were correctly named as the name of the painting presented in the photograph. From the 10 uncorrectly classified photographs 5 (3 in SURF) was falsely classified as not present in the database and 5 (7 in SURF) was classified with the name of the wrong painting. Unlike the simmilar classification results, the SUFT method prove to be 2 times faster in the creation of the descriptor files. The best segmentation measure was 89 % and the classification measure was 90 %.

8 Future work

In the next phase of the work the classification process will be improved by adding additional criterion to the matching process. The paintings will be compared also by the aspect ratio. This criterion will be helpful for the images with the best matching value close to the threshold. This value may also accelerate the matching process- the descriptor file of the image will be matched only with DFs of the paintings with the similar ratio values.

9 Acknowledgments

References

[1] Inc. AG.com. Travel webshots., 2009. http://travel.webshots.com.

[2] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346 – 359, 2008. Similarity Matching in Computer Vision and Multimedia.

[3] Gertjan J. Burghouts and Jan-Mark Geusebroek. Performance evaluation of local colour invariants. *Comput. Vis. Image Underst.*, 113(1):48–62, 2009.

[4] Etezadi-Amoli M. Chang C. and Hewlett M. A day at the museum., 2009. http://www.stanford.edu/class/ee368/Project07/reports/ee368group06.pdf.

[5] Florin Cutzu, Riad Hammoud, and Alex Leykin. Distinguishing paintings from photographs. *Computer Vision and Image Understanding*, 100(3):249 – 273, 2005.

[6] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.*, 40(2):1–60, 2008.

[7] Xiong G. Gradient using first order derivative of gaussian., 2009. http://www.mathworks.com/matlabcentral/fileexchange/8060-gradient-using-first-order-derivative-of-gaussian.

[8] Shuqiang Jiang, Qingming Huang, Qixiang Ye, and Wen Gao. An effective method to detect and categorize digitized traditional chinese paintings. *Pattern Recognition Letters*, 27(7):734 – 746, 2006.

[9] Thomas Edward Lombardi. *The classification of style in fine-art painting*. PhD thesis, New York, NY, USA, 2005. Adviser-Tappert, Charles and Adviser-Cha, Sung-Hyuk.

[10] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2002.

[11] Jiri Militky. Image analysis and matlab., 2009. http://centrum.tul.cz/centrum/itsapt/Summer2005/files/militky_6.pdf.

[12] Mataev Olga. Olga's gallery., 2009. http://www.abcgallery.com/index.html.

[13] Pietro Perona and Jitendra Malik. Scale-space and edge detection using anisotropic diffusion. Technical report, Berkeley, CA, USA, 1988.

[14] Hough P.V.C. Method and means for recognizing complex patterns.

[15] Robert Sablatnig, Paul Kammerer, and Ernestine Zolda. Hierarchical classification of paintings using face- and brush stroke models. In *in Proc. of 14th Int. Conf. on Pattern Recognition*, pages 172–174, 1998.