

Graphism Recognition in Goya's Work

Amparo Sacristán Carrasco
Polytechnics School of Mataró
Dept. Telematics and Computer Architecture
Avda. Puig I Cadafalch 101-111
08303, Mataró, Spain
sacrista@eupmt.cat

ABSTRACT

In recent years many experimental specialists have carried out analysis of some pictures to test if they are authentic works of art. This work can help expert people to authenticate that kind of pictures. This paper describes the development of a fully-automated computerized analysis system of Goya's pictures, which enables to locate some small prints (graphisms) in Goya's authentic work. The proposed methodology works on four steps: digital image acquisition on gray level, segmentation, pattern recognition and localization of small prints in the picture. The outcome of the process is the location of some small prints (graphisms) if the input digital image is a Goya's picture. An artificial backpropagation neural network is used to say if graphism segmented is a Goya's small print or no.

Keywords

Image segmentation, pattern recognition, artificial neural network, analysis system

1. INTRODUCTION

Goya is one of the greatest painters of 18th century. Goya taught a lot of students his technique and style of painting and drawing in his school, so there are a lot of works that are not authentic of Goya but they can be catalogued as Goya's, because they were made by their disciples. Also, some contemporary painters imitated and falsified Goya's work [Mor94]. Goya constantly experimented on new techniques and broke his own rules, developing a very sharp style [Sch07]. He mixed portraits, self-portraits, landscapes, religious themes and others. So the work of the expert specialists is very hard as they must study a lot of parameters in each work of art.

After studying a great number of Goya's work, it is possible to claim the next hypothesis [Veg91]: "*Goya carried out his work with characteristic strokes, made up with spelling of his surname (G – o – y – a). He shaped to lines, outlines, light touches, shadows, contrasts, ... always being covered up*".

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Copyright UNION Agency – Science Press, Plzen, Czech Republic.

These small prints are called *graphisms*. Different authors have shown those small prints in Goya's work ([Med70], [Fau96]) as particular characteristics of painter. But since the works of [Agu97] and [Rod99] the extraction of this characteristic was not automated. Another work [Veg00] was previously made but this one was not automatic.

This paper describes an improvement on the automation of graphisms localization. And it is possible to verify the process of pattern recognition at the same time as artificial backpropagation neural networkwork is testing the possible graphisms.

Graphisms are difficult to see to the naked eye, if the person who is looking for them does not have experience in this kind of information. But, this problem can be solved with digital image processing. The goal of this work is to locate graphisms automatically to make their visualization possible to non-expert people.

Once verified that graphisms are in all kind of Goya's work, it was necessary to highlight them. The process used to stand out graphisms is based on computer-based digital image processing, which enables to take out information from the image although it is not visible to human eye.

This paper starts by giving the requirements for acquisition of digital image in section 2. Section 3 presents the algorithms to extract the graphisms. The description of all this application is in section 4.

Some results are shown in section 5. And several conclusions are provided in section 6.

2. IMAGE ACQUISITION

Original images may be a very important and expensive material so this work is based on digital photographs ones. The photograph can be analog in DIN-A4 size.

It was necessary to balance between resolution and quality at the time of image acquisition. After different tests with images and testing positive results, the decision was made to digitalize the images with a resolution of 100 or 300 ppi depending on input image size. Low resolution digitized images can distinguish a big number of graphisms with different intensity, crossed with others, superimposed, etc., if eyes looking them have experience on this kind of information. But this class of graphisms has a great difficulty to be processed.

The digital image is gray level, not a true color, because we are looking for line information. That kind of strokes are the same in color or gray and the process will be faster by using gray level code, so the storage memory will be smaller.

We use BMP image format to put digital image in storage because information can be lost if compression is used.

3. EXTRACTION OF GRAPHISMS

The extraction of graphism pixels is very difficult because of their characteristics. Sometimes there are very small differences between the gray levels of pixels in a graphism and their neighbors. Due to that complexity, the extractions of graphisms from digitalized images are made by using two different algorithms:

- *Region segmentation*
- *Graphisms segmentation*

First, an interval is defined in the range of 16, 32 or 64. Then, a region is defined as set of pixels with more or less the same gray level. The maximum difference of gray level between pixels of the same region is the interval value.

The next sections give an account of *Region Segmentation* and *Graphism Segmentation* algorithms.

3.1 Region Segmentation

This algorithm segment the original image into regions R.

The *Region Segmentation Algorithm* is divided in three steps:

Step1: Eliminate possible noise applying a median algorithm.

Step2: Classify all gray levels I.

Step3: Each gray level is segmented in different regions following the steps:

3.1. P is a pixel with a gray level L

3.2. Define $j = \frac{i}{2}$, where i = 16, 32 or 64

3.3. R is the image with all of pixels Q with gray level in the range $[L - j, L + j]$

The step1 is necessary because noise can be introduced during image acquisition or which is on the original art picture.

The steps 3.2 and 3.3 are repeated while there are non-processed pixels and verifying the condition $[L - j, L + j]$.

The Figure 2 shows an example of region of 'The Drinker', Goya's oil painting (Figure 1)



Figure 1. 'The Drinker' digitalized with 256 gray levels



Figure 2. Region from 'The Drinker' with 32 as value interval

Different number of regions can be segmented into the same intensity I of an interval. Each region is defined from a different pixel with intensity L. This pixel must not be processed in other region.

3.2 Graphism Segmentation

This algorithm segments a region into areas of connected pixels or pixels that are at a distance of 3 pixels as maximum. A distance of 3 as maximum is considered because graphism can be painted in a color and afterword it will be painted in another color crossing it with other colors. Ten part of the graphism will be in other region.

This algorithm is divided into two steps:

Step1: Areas are segmented from a region R:

1. While there are pixels in R
 - 1.1. $P \in R \Rightarrow P \in A_i$, where A_i is the area i
 - 1.2. While there are pixels Q in region R:
 - if $d(Q, A_i) \leq 3 \Rightarrow Q \in A_i$
 - else turn over step 1.1 with $P=Q$.
- where d is the distance between the pixel Q and region A_i

Step2: Each area A_i is scaled to have the size of standard input of classifier.

- 2.1. [m,n] are dimensions of area A_i
- 2.2. if $m \geq n \Rightarrow m = 22$
else $n = 22$
- 2.3. The other dimension is calculated to preserve the ratio of A_i dimensions

It is clear there will be a lot of areas which will not be graphisms. A small number of them will contain this kind of information.

Table 1 shows a part of HTML file, which is one of the outputs of *Graphisms Segmentation*.




	regio1_GRAFISME13.bmp	
	(27,165)	(30,165)
	(27,168)	(30,168)
	regio1_GRAFISME14.bmp	
	(30,129)	(35,129)
	(30,135)	(35,135)
	regio1_GRAFISME15.bmp	
	(35,146)	(41,146)
	(35,152)	(41,152)

Table 1. Areas segmented with their coordinates

A scaled area is shown in figure 3. Areas are scaled at 22 pixels at largest dimension, and the ratio is preserved. The scaled areas are binary images. The

pixels of an scaled area has the value 0 so it will be necessary to make the negative of that image before it will be shown as input at the classifier.

The classifier is an artificial neural network of the backpropagation kind.

```

1111111111111111000
1111111111111111011
1111111111111111011
1111111111111000000
1111111111111000011
1111111111111000011
1111110111110000000
1111111001000111111
1111111001000111111
1111111111000011111
1111111000111111111
1111111000111111111
1111110000111111111
1111000111111111111
1111000111111111111
1110000111111111111
1000111111111111111
1000111111111111111
0000111111111111111
0000111111111111111
0000111111111111111
0000111111111111111

```

Figure 3. Area segmented as possible graphism

4. GRAPHISM RECOGNITION

This is an application developed with C++ Builder 5.

The proposed methodology works on four steps:

Step1: Segmentation of digital image.

- 1.1. Region Segmentation
- 1.2. Graphisms Segmentation

Step2: Classification of areas in graphisms (section 4.1).

Step3: Localization of graphisms and statistics (section 4.2).

The segmentation step has been presented in the section before.

4.1 Classification

The classifier is a backpropagation artificial neural network with 484 neurons at the input level; two hidden levels with 50 neurons each and an output level with 5 neurons. The output level has 5 neurons because the fifth neuron is assigned to areas which are not graphisms (figure 4 shows an authentic graphism and figure 5 shows a segmented area that is not a graphism). Different kind of architectures were tested and the best successes were presented with this one.

This part of application has different possibilities:

1. Train the classifier
2. Classify the scaled areas as graphisms G, o, y, and a.

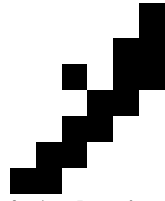


Figure 4: Authentic graphism (original size)

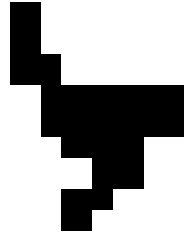


Figure 5: Segmented area (no graphism) with original size

4.1.1 Train the classifier

This artificial neural network was trained with different sets of patterns (figure 7), all of them are small prints of Goya's authentic signatures (figure 6) [Can81]. A subset of small prints was selected, being the same number for each kind of print. Each kind of prints (G, o, y and a) was sorted according to their correlation before being shown to backpropagation artificial neural network. Also different orders were considered for each of types of prints .

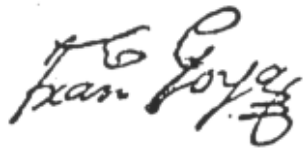


Figure 6: Authentic Goya's signature

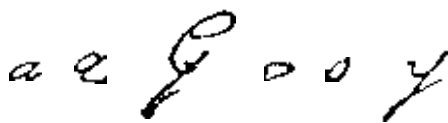


Figure 7: Prints of Goya's signature

Training patterns shown in groups with the same order as the painter's surname (Goya) gave the best successes at the classification time.

The generated weights at the end of the training process must be saved. They must be loaded every time one set of segmented areas need to be classified.

4.1.2 Classifying scaled areas

The verification process can be manual or automatic. A manual verification offers the possibility to know the kind of graphism at each time. While automatic process only allows to know the classification at the end by looking through the generated file along the verification process. This file is called STATS. The stimulated output and its value are written in STATS file, and this kind of information is used in the localization process to make statistics.

A graphism classification is shown in table 2, the segmented area is on the left. To the middle there is scaled area and the classification is on the right.

	<pre> 1111111111111111000 1111111111111111011 1111111111111111011 1111111111111111011 1111111111110000000 1111111111110000111 1111111111110000111 1111111111110000111 1111111111110000000 1111111001000111111 1111111001000111111 1111111000111111111 1111111000111111111 1111110000111111111 1111000111111111111 1111000111111111111 1110000111111111111 1000111111111111111 1000111111111111111 0000111111111111111 0000111111111111111 0000111111111111111 0000111111111111111 </pre>	<p>2 → 0.9859</p>
--	--	--------------------------

Table 2. Step to step of a graphism classification

The stimulated output neural network coding is

- 0 means graphism 'G'
- 1 means graphism 'O'
- 2 means graphism 'Y'
- 3 means graphism 'A'

Then table 2 shows a segmented area (left column) which is classified as 'Y' graphism and the output neural network is 0.9859 (right column).

4.2 LOCALIZATION

It is possible to localize the graphisms over the original image or over a region. The localization over the original image makes the visualization more difficult, because graphisms can be in lines or intensities not visible to the naked eye. It is better to do the localization over regions. They are binary images and that makes the visualization easier.

The process of localization makes the two localizations at the same time: over original image and over regions. At the same time, statistics are shown. The summary of statistics can refer to the graphism classification in a full image, in a gray level or for an interval.

Localization of graphisms for region of figure 2 is shown in figures 8 and 9.

A little signature is shown in the left bottom of figure 8. There are a lot of lines that look like graphisms but they are not emphasized with a square, because the stimulated output neural network was smaller 0.95.

Whereas graphisms localization over original image is shown in figure 9.

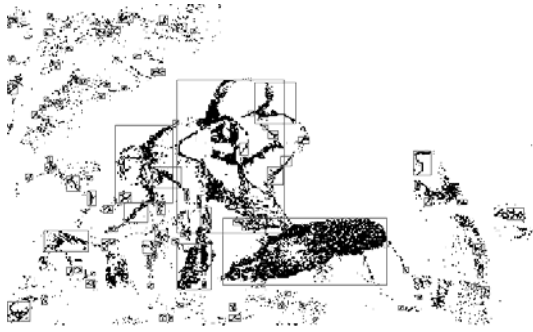


Figure 8. Localization of graphisms over a region

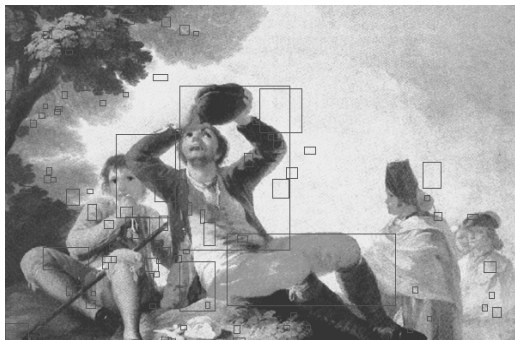


Figure 9. Graphisms localization over original image

Figure 10 shows the statistics for localization displayed in figures 8 and 9.

- Grafismes del tipus G: 19
- Grafismes del tipus O: 9
- Grafismes del tipus Y: 13
- Grafismes del tipus A: 14

Figure 10. Statistics for a region

5. RESULTS

To evaluate the efficiency of our application:

1. We looked through the segmented regions to verify if there is a real graphisms on the coordinate which was automatically reported.
2. The same graphism was verified over the original image.

This process to verify the correct localization of graphisms was done with Goya's work and with works of other painters as Velazquez, Dalí, Picasso and Fortuny.

This application was tested with Goya's

- oil paintings (figure 11),
- drawings (figure 12)
- engravings (figures 13, 14 and 15).

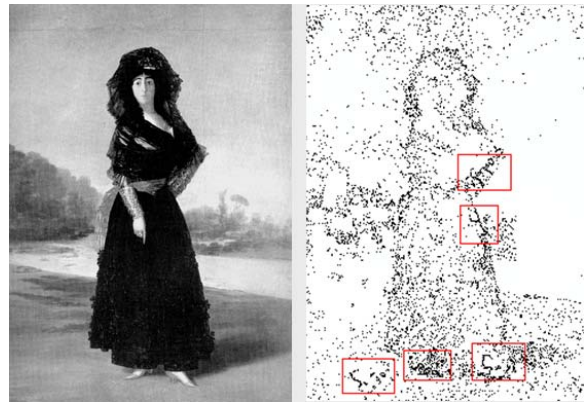


Figure 11: Segmented Region with localization of graphisms



Figure 12: Segmented Region with localization of graphisms

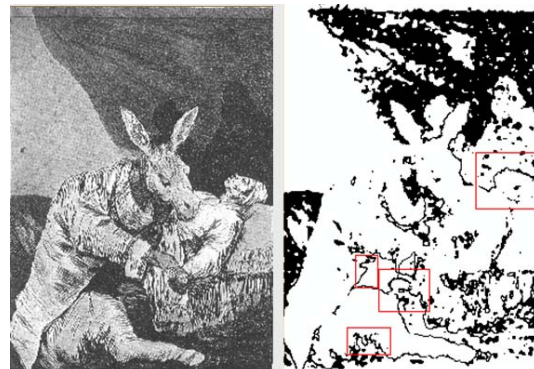


Figure 13: Segmented Region with localization of graphisms (on the left 'From that evil die?' original Goya's engraving)



Figure 14: Segmented Region with localization of graphisms (on the left 'From that evil die?' original Goya's engraving)

Global Goya's work is catalogued into four periods. Four drawings of Goya's first period were selected

and two oil paintings of the second, third and fourth periods. About engravings, we chosen two engravings of each group: Whims, War Disasters, Bullfighting and Pieces of Nonsense.



Figure 15: Segmented Region with localization of graphisms (on the left segmented region of 'Populacho')

The application localized around 1000 correct graphisms in full original image, but we only shown in this paper graphisms localized on segmented regions because they are easier to watch.

The ratio of *false positive*¹ (FP) and true positive (TP) are the same for each kind of Goya's work, it is over 0.27 in full original image and 0.15 over regions. We compare these values with previous versions ones and they are better in this case. And we have derived better ones changing the set of training patterns.

When we tested the work of other painters the number of possible graphism was very small, and never a 'G' or 'Y' were classified. When we verified these results over the region and original images, we were not able to find them so all of them were false positive.

6. CONCLUSIONS

The application presented in this paper is a new contribution to help experts in art to catalog Goya's work. Experts in art use more and more computers to catalog works of art correctly [Ver07].

The localization of graphisms is made in a fully automatic way, thus making subjective facts disappear.

Some authors as [Med73], [Veg91] said there are a lot of graphisms in Goya's work, and that was proven with Aguilera's work [Agu97]. But Medgessy in his investigation [Med73] visualized them by radiography of canvas and De la Vega noticed presence of graphisms making a video over Goya's work [Veg91]. The last one uses Roure's Method [Veg00] to highlight the graphisms, but that is a method with manual work and it can present subjective view.

¹ *False positive* are graphisms segmented by the application but they are not authentic Goya's prints

This work does not intend to look for signatures as other authors have done. It only looks for small prints present in Goya's work ([Agu97], [Rod99]). Other painters' works of art were processed with this application but graphisms were not localized. In general, all of segmented areas were classified as no-graphisms, only in few cases the Graphisms Recognition localized some lines as graphisms 'A' and 'O', which are not significant.

Other painters' works of art (Velazquez, Dalí, Picasso and Fortuny) have also been analyzed to prove that lines as graphisms were not unwilling painted in Goya's..

The size of segmented areas must be higher 3x3 pixels and very long and narrow lines are not segmented.

In the future, it will be necessary to find a better pattern set to train the classifier and to be able to minimize the number of false positive. The number of FP in a full image is around 20%, this number is smaller when the localization is made over regions, which it is less than 10% in this case.

7. ACKNOWLEDGMENTS

Our thanks to Mr. Xavier Amir and Dr. Salvi Sendra to pay for part of this work, and to Mr. Jordi Roure (R.I.P.) who presented the problem of localization graphisms at Politechnics School of Mataró.

8. REFERENCES

- [Agu97] Aguilera, N. and Sacristán, A. "Automatización del proceso de obtención de grafismos en la obra de Goya" in URSI'97 conf. proc. Vol. II, pp. 197-199. IBERDROLA INST. TEC. Press I.S.B.N. 84-89654-03-4, Bilbao (Spain)
- [Can81] Canellas, A. Diplomatario de Francisco de Goya. Ed. Á. Canellas. Zaragoza, 1981.
- [Fau96] Faúndez, M., Sacristán, A., Espinosa, V. and Caballero, M."Generación de un OCR para el reconocimiento de grafismos ortográficos y soporte a la autentificación de cuadros de Goya" in URSI'96 conf. proc. Vol. II, pp.209-212. Servicio de publicaciones E.T.S.I. Telecomunicación. D.L. M-27551-1996, Madrid (Spain)
- [Med70] Meddassy, R. "El arte perdido de Goya". Revista La Actualidad Española". Abril 1970
- [Mor94] Morales y Marin, J.L. "Goya. Catálogo de la pintura". Alpuerto S.A. Press 1994
- [Rod99] Rodríguez, S. Hernández, J.M. Arévalo, R. and Sacristán, A. "Reconocimiento de grafismos ortográficos en la obra deGoya" in URSI'99 conf. proc. pp.195-196.I.S.B.N. 84-699-0893-6. Santiago de Compostela (Spain)
- [Sch07] Schulz, a. Aesthetics, Perception and the Body. Cambridge University Press, 2007. ISBN-13: 9780521821056 | ISBN-10: 0521821053
- [Veg91] Vega, J.I.de La "Los rostros ocultos de Goya". Congreso de Bellas artes. Junio 1991, Barcelona (Spain)

[Veg00] Vega, J.I.de La “GOYA (II.Roura’s Method)
<http://www.goya2000.com>

[Ver07] Veritus AC. Digital Verification Technology
<http://www.veritus.org>