Inkjet Printing discrimination based on invariant moments

Vanessa Talbot, Patrick Perrot, Cyril Murie
Institut de Recherche Criminelle de la Gendarmerie Nationale,
1, boulevard Théophile Sueur
93111 Rosny sous bois cedex
France

Abstract

In the field of forensic science the question of finding out a solution to discriminate ink jet printings, provides interesting indicators to investigators. Nevertheless, this task is not obvious because of several parameters such as the media type, the artefact of the printer, the age of the printer and so on. The aim of this study is to identify automatically type and model of a printer from the characters of an anonymous letter for instance. Generally, it is not possible to distinguish a printer from another, just by a visual inspection of the writting. Our work based on recognition pattern consists in finding some features extracted from the letter « a », able to characterize the printer. A stochastic approach is used to identify the invariant features of the letter « a » printed by three kinds of printers: Epson Stylus, Canon i905, HP Photosmart. The principle of the method is based on the calculation of seven invariant moments proposed by Hu[3]. The distribution of each moment is modelled by a gaussian component in a training phase which contains 80 letters. The test phase is based on different conditions. The first one consists in identifying a printer. The second one evaluates the influence of three word processing software, and at last, the third one proposes a study of the scanner effect. The obtained results reveal that printer discrimination is possible independently from the word processing software. And last the scanner effect decreases significantly the power of discrimination according to the resolution used.

Introduction

Printed material is very used for criminal activities or terrorism: anonymous threatening letters, terrorism claim and so on. The main advantage is that a printed letter is more anonymous than a written letter. The challenge for investigators is to be able to find the authors of the letter and all kinds of elements could be determining. The idea of identifying the device used to print the document is very interesting, but not very easy to achieve.

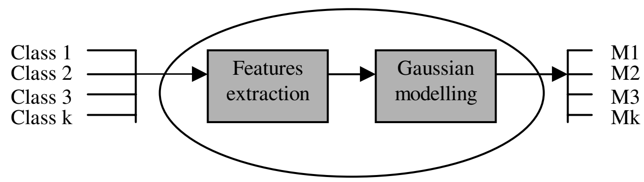
So, our aim in this work is to propose a technique which identifying printers (model and type) automatically independently from the ink. There are two main categories of home office printer: laser printers (xerography) and inkjet printers. Our study is limited to this last group. The aim is to determine a specific signature for each printer and not a “printerprint”.

Our method of classification is closely based on pattern identification. The idea is to extract specific features which characterize the printer and classify them from a Gaussian distribution.

So, we present in a first section an overview of our technique before to detail in a second one the features extraction. The modelling technique and the experimental results will follow in the third and fourth section.

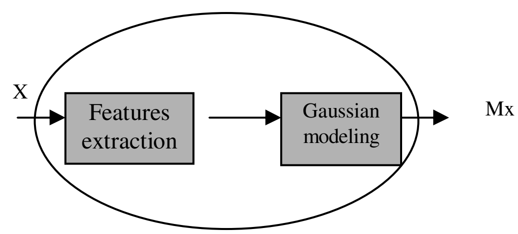
Overview of the system

The identification system is based on two steps: training and testing modules. Figure 1 and Figure 2 shows the block diagram of the identification system.



Class1: Printer Group n°1
M1: model for class 1

Figure 1: training phase



X: unknown document
Mx: Model of X

Figure 2: Testing phase

From a character database for each kind of printer, we extract some specific parameters which the distribution is represented by a Gaussian. Given a anonymous letter, the same features extraction and modeling phases are carried out and a visual comparison between the distribution model of the unknown document and the distribution of the printer model provides some interesting elements of discrimination.

The study focuses on the character "a" which has been extracted on the different documents. We chose the "a" size 12 (Times and Times New Roman) because of the geometrical characteristics of this character. This geometry is well-adapted to the feature extraction. The size 12 has been chosen because this is the most frequently occurring in documents.

Epson stylus Canon i905 Hp photosmart

Experiments were carried out on three kinds of printers:

- Epson stylus,
- Canon i905,
- Hp photosmart.

These manufacturers represent the most common printers used. For these different printers, we extracted 80 characters to build the model: one model for one class of printer.
So the aim of the system consists in proposing if a unknown document was printed by the Epson stylus, the Canon i905 or the HP photosmart from this analysis of specific features on the character "a".

Feature Extraction

The first step of this work previously presented, is to extract some relevant features in order to discriminate the different printers. The features proposed in our study are the Hu moments[3]. Hu described a set of seven moments that are rotation, scaling, translation and skew invariant. We can find other invariant moments in the literature[1]: Legendre moments or Complex Zernike moments[4]. These are not studied in this paper.

The Hu moments are based on a non-linear combination of normalized moments. These features are particularly adapted to the description of a picture but not to its reconstruction. So these moments are extracted for each character "a".

Considering I(x,y) as the pixel grayscale level of the picture I. The moment of a picture I for the order p+q (p,q>0) is defined like:

$$m_{p,q} = \int_{R^2} x^p y^q I(x,y)\,dxdy$$

The centroïd of the I function is: $x_0 = \frac{m_{1,0}}{m_{0,0}}$; $y_0 = \frac{m_{0,1}}{m_{0,0}}$

The centered picture I_T ($x+x_0$, $y+y_0$) is translation invariant. The central moment of the I picture for the order p+q is:

$$v_{p,q} = \int_{R^2} x^p y^q I(x+x_0, y+y_0)\,dxdy$$

These last moments are translation invariant. And the normalized moment are defined like:

$$\mu_{p,q} = \frac{v_{p,q}}{v_{0,0}^{1+(p+q)/2}}$$

These moments are translation and scaling invariant. The Hu moments are computed from these normalized moments and they are translation, rotation and scaling invariant as previously said.

Mh1 = $\mu_{2,0} + \mu_{0,2}$

Mh2 = $(\mu_{2,0} - \mu_{0,2})^2 + 4\,\mu^2_{1,1}$

Mh3 = $(\mu_{3,0} - 3\mu_{1,2})^2 + (3\mu_{2,1} - \mu_{0,3})^2$

Mh4 = $(\mu_{3,0} + \mu_{1,2})^2 + (\mu_{2,1} + \mu_{0,3})^2$

Mh5=$(\mu_{3,0} - 3\mu_{1,2})(\mu_{3,0} + \mu_{1,2})[(\mu_{3,0} + \mu_{1,2})^2 - 3(\mu_{2,1} + \mu_{0,3})^2] + (3\mu_{2,1} - \mu_{0,3})(\mu_{2,1} - \mu_{0,3})[3(\mu_{3,0} + \mu_{1,2})^2 - (\mu_{2,1} - \mu_{0,3})^2]$

Mh6 = $(\mu_{2,0} - \mu_{0,2})[(\mu_{3,0} + \mu_{1,2})^2 - (\mu_{2,1} + \mu_{0,3})^2] + 4\mu_{1,1}(\mu_{3,0} + \mu_{1,2})(\mu_{2,1} + \mu_{0,3})$

Mh7 = $(3\mu_{2,1} - \mu_{0,3})(\mu_{3,0} + \mu_{1,2})[(\mu_{3,0} + \mu_{1,2})^2 - 3(\mu_{2,1} + \mu_{0,3})^2] - (\mu_{3,0} - 3\mu_{1,2})(\mu_{2,1} + \mu_{0,3})[3(\mu_{3,0} + \mu_{1,2})^2 - (\mu_{2,1} + \mu_{0,3})^2$

As described in figure 1 and 2 the features extraction regards the training and the testing phases. In the first case for each printer we extract from a document the 80 character "a" and compute the Hu moments. So we obtain a set of 7 moments for each class of printer. In the second case (testing phase) the same operation is achieved to elaborate a set of seven moment for the "a" of the unknown document.

Printer Modeling

From the set of features we represent the distribution of these parameters by a Gaussian distribution. The distribution of each feature for each printer is represented in the following figures:

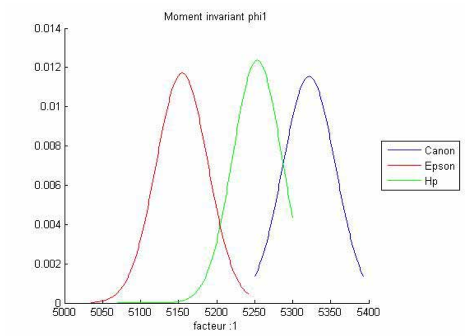


Figure 3: Mh1 for the three printers

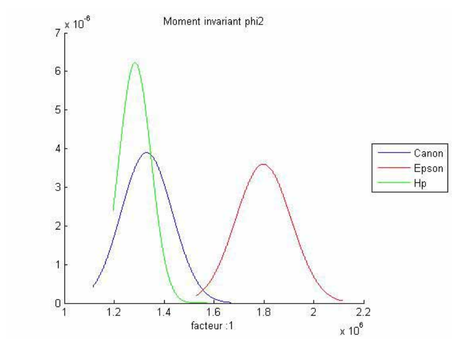


Figure 4: Mh2 for the three printers

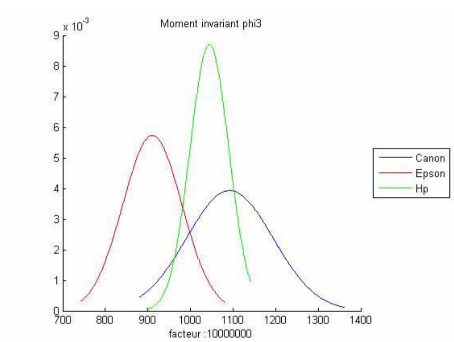


Figure 5: Mh3 for the three printer

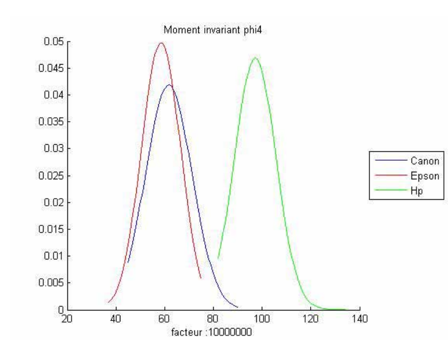


Figure 6: Mh4 for the three printers

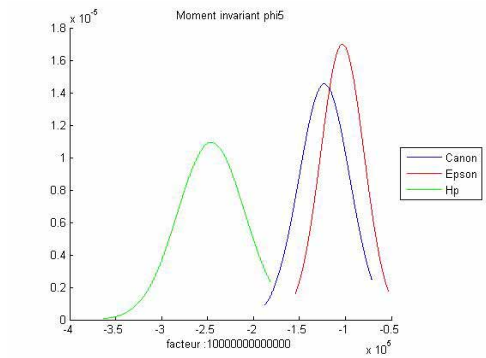


Figure 7: Mh5 for the three printers

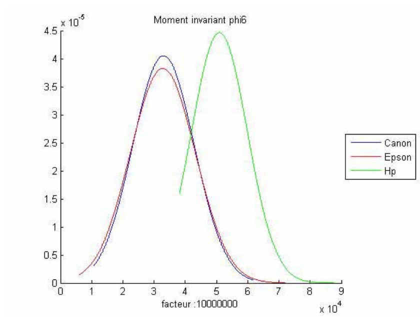


Figure 8: Mh6 for the three printers

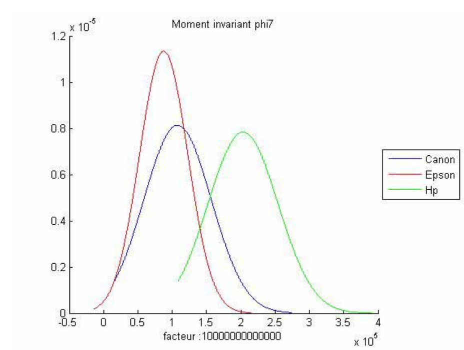


Figure 9: Mh7 for the three printers

The analysis of this distribution provides some interesting possibilities of discrimination

Experimental results

Three kinds of test have been achieved in order to study:

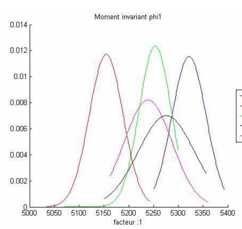
- test n°1: the power of the Hu moments for Printer discrimination
- test n°2: the influence of the word processing software
- test n°3: the influence of the scan quality

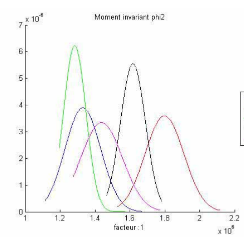
➢ Test n°1:

The previous distribution presented is used to identify a set of testing printers composed by:

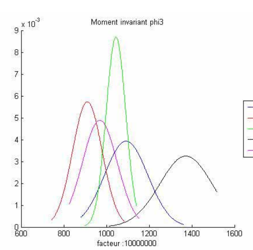
- Hp printer called test 2
- Unknown printer (not in the training database) called test 1

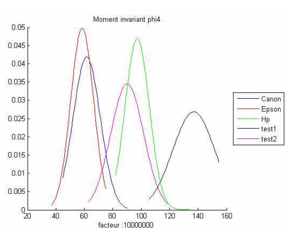
The results are significant because a comparison of the distribution provides some relevant elements of similarities between the test Hp printer and the class of the Hp printers contrary to the unknown printer which the distribution is very different for several moments (Mh3, Mh4, Mh5, Mh6, Mh7)



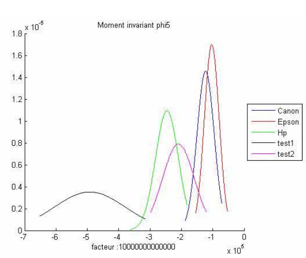


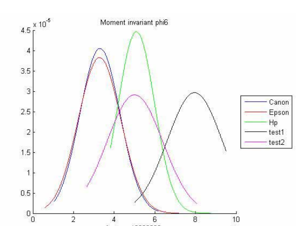
Mh1 Mh2



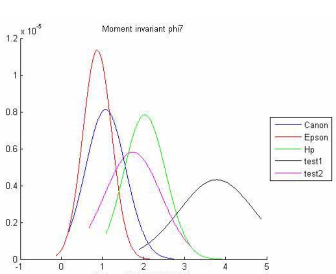


Mh3 Mh4





Mh5 Mh6



Mh7

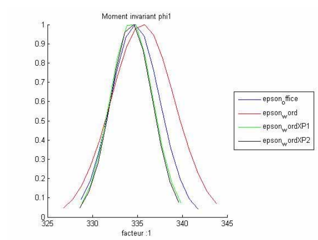
Figure 10: Printer comparison based on Hu moments

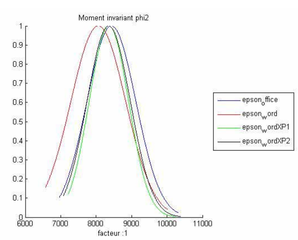
➢ Test n°2

A complementary experiment has been carried out in order to evaluate the word processing software:

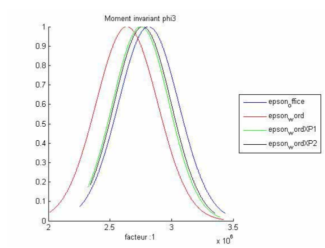
- Office 1.1.4
- Word 2002
- Word XP1 (1 scan)
- Word XP2 (after 2 scan)

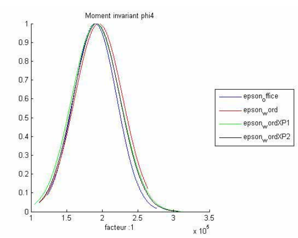
The results tend to prove that there is no difference caused by a different word processing used. So, the discrimination of printer based on Hu moments is independent from the word processing software. And in addition we do not notice any change between the two distributions of Word XP that is to say that there is no influence of the number of scan on the power of discrimination. The results are presented in the following figures.



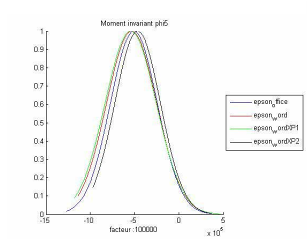


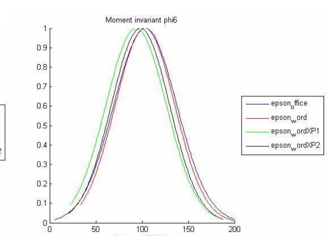
Mh1 Mh2



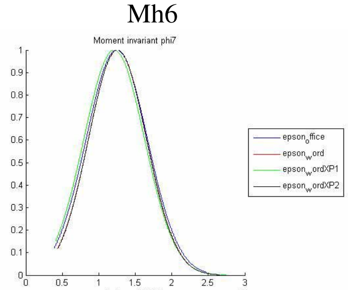


Mh3 Mh4





Mh5 Mh6



Mh7

Figure 11: Results on Hu moments

- Test n°3

Four printers have been used to study the influence of the scan quality in the power of discrimination: Hp photosmart, Hp deskjet, Canoni905 and Epson stylus. A decreasing of the scan resolution degrades significantly the power of discrimination of the printers. So the quality of the picture is important to get some relevant results.

Conclusion and Perspectives

Our experiments were carried out on a limited set of data because we have just used 80 characters to evaluate the distribution of the features in the training phase for each group of printer and we have not used lots of printers for test. Nevertheless the results obtained provide some interesting ways for discriminating printers.
Future improvements should be planned by computing a distance between the distributions based on the power of discrimination of a GMM (Gaussian Mixture Models) [2] or a SVM (Support Vector Machines) classifer for instance and by testing some others invariant moments as proposed in [1][4][5].

References

[1]. Choksuriwong, A.; Laurent, H.; Emile, B. *Comparison of invariant descriptors for object recognition*, Image Processing, 2005. ICIP 2005. IEEE International Conference on volume 1, Issue , 11-14 Sept. 2005 Page(s): I - 377-80

[2]. Gazi N. A., Aravind K. Mikkilineni, Edward J. Delp, and Jan P. Allebach, *Application of Principal Components Analysis and Gaussian Mixture Models to Printer Identification*, Proc. IS&T's NIP20, International Conference on Digital Printing Technologies 2004

[3]. M. Hu, *Visual pattern recognition by moment invariant* , IRE Trans on Information Theory, Feb 1962.

[4]. A. Khotanzad, Y. Hua Hong. *Invariant image recognition by Zernike moments.* IEEE transactions on Pattern Analysis and Machine Intelligence, 12(5):489-497, 1990

[5]. M. Teague. *Image analysis via the general theory of moments.* Journal Optical Society of America, 70:920-930, 1980

Author Biography

Vanessa Talbot received her Master's Degree in Electronic and Optic from Institut Supérieur des Techniques Avancées, Saint-Etienne (ISTASE), France, in 2003. Since 2004 she has worked at IRCGN (Forensic Research Institute of French Gendarmerie) in Rosny sous Bois - France - in the Audio, Video and Signal Processing department. Her main researches focus on finding the characterization and the modeling of speech features. Her work in the field of printer characterization results from a collaboration with the IRCGN Document department