Comparison between Pen-scanner and Digital Camera Acquisition for Engraved Character Recognition

Céline Mancas-Thillou
Faculté Polytechnique de Mons
Belgium
celine.thillou@fpms.ac.be

Matei Mancas
Faculté Polytechnique de Mons
Belgium
matei.mancas@fpms.ac.be

Abstract

Engraved characters are usually a limitation for character recognition, as their extraction is difficult due to the small contrast with the background. Nevertheless, they are present in several industrial applications and their handling may be convenient for special tasks such as traceability or quality control. We propose here an entire system from acquisition to character recognition by considering two different and mobile acquisition devices, a pen-scanner close to traditional acquisition and a digital camera in order to appreciate performance and differences. In some industrial applications such as the ones including engraved characters, it may be difficult and awkward to properly build a large training database for recognition. Hence, we also propose a novel method to increase automatically the size of the training dataset without loosing representativeness by using image analogies. Results are finally presented in terms of Precision and Recall for character extraction quality and recognition rates for character recognition with comparison with a commercial OCR.

1. Introduction

Machine vision and image processing algorithms have attempted for several years to be used in industrial applications. With the drastic expansion of digital cameras with reasonable prices, the particular domain of pattern recognition offers promising applications for industry. A gap needs to be filled in order to make image-based methods more efficient and versatile. One of the existing limits, nowadays, is for engraved character recognition. As stated in [13], camera-based techniques for character understanding give lower recognition rates for engraved characters. In industry, those patterns are quite present, especially on metallic objects with registration numbers, for example. Hence additionally to the engraved issue, these surfaces are reflective, which is still a topic in research in order to build robust algorithms against uneven lighting. The challenge is double: efficient understanding of engraved characters on reflective surfaces and training database constitution. Both points will be addressed in this paper.

For the first issue, a convenient mobile acquisition is needed. For several reasons of portability, robust and light devices are often required. Moreover, engraved characters are part of objects, which are very rarely flat. Hence, usual scanner-based acquisition can not be a solution. Contrarily to traditional scanners, pen-scanners offer mobility, handling of non-paper objects in every context and may constitute a real choice for industrial applications. Similarly, digital cameras give the same benefits and have been extensively used in several industrial applications such as tourist dedicated translation systems [25] or reading devices for visually impaired [14]. Video-based systems have been tested in this context and are not competitive in terms of results quality for acceptable prices. Heavy processing using super-resolution algorithms is required and leads to low-quality results compared to pen-scanners and still cameras. For these reasons, a comparison between these two devices will be addressed in the following sections.

The second issue of training database constitution for character recognition is a common problem in all pattern recognition algorithms. However, it is even more highlighted with engraved characters present in industrial situations. It is difficult to acquire large databases with various examples. No dedicated training for engraved characters lead to poor results. Hence a solution for the artificial increase of training database is absolutely needed.

Character extraction and recognition in the context of engraved character present several applications, such as quality control or traceability. Numbers need to be clearly engraved without ambiguity for end-users, especially in the case of use-before date on cans or for inspection at different creation times for risky objects, such as weapons (Figure 1). Character recognition is an interesting alternative for unalterable objects. Several other situations may also be listed.

Figure 1. A registration number on a weapon butt (Copyright FN Herstal, Belgium, 2007).

Section 2 will be referred to the state-of-the-art of character extraction, useful for engraved characters. Section 3 will be presented our methods in terms of character extraction and uneven lighting removal for both mobile acquisition devices. In Section 4, our training database constitution and artificial increase will be presented with our inhouse character recognition. Section 5 will be addressed to detailed results in terms of character extraction and recognition. Comparisons with existing algorithms will also be mentioned. Finally, in Section 6, a discussion will conclude this paper along with presentation of our future works.

2. State-of-the-Art

As far as we know, no systems have been dedicated to engraved or embossed characters, mainly due to the recent work on camera-based systems and their analysis complexity. Recent color-based text extraction can not efficiently use color information [13] for this kind of characters, present in a metallic environment. The main point is the close similarity between color foreground and background. Moreover, due to the metallic environment, colors are close to the main RGB diagonal, meaning they mostly represent variations of gray. Hence, these methods compete with binarization systems from gray-level images.

Natural scene character extraction algorithms are classified into several categories, from the simplest ones to the most difficult ones:

Thresholding-based methods define a threshold globally (for the whole image) or locally (for some given regions) to separate text from background. Histogram-based thresholding is one of the most widely used techniques for monochrome image segmentation. The threshold is chosen as the value corresponding to the valley between two peaks. The most referenced method is the one described by Otsu [17] and used for a visually impaired-driven application in [3, 21] . These methods work well with low computational resources and are applied mostly on gray-scale images. Adaptive or local binarization techniques define several thresholds $T(i, j)$ for different image parts depending upon the local image characteristics. Several papers [11, 24] for video text extraction used the Niblack's method [16] where the threshold depends on local mean and standard deviation over a square window of size to define. An extension is the method of Sauvola and Pietikäinen [19]. This adaptive technique is in use in *Mobile Reader*^{TM} [9], a mobile phone reading text from Inzisoft. Adaptive binarizations may handle more degradations (uneven lighting, varying colors) than global ones but suffer to be too parametric which is not versatile. Moreover, these techniques still consider gray-scale images only and were mainly used for video caption text with clean backgrounds.

Entropy-based methods use the entropy of the gray levels distribution in a scene. Li and Doermann [11] minimized the cross-entropy between the input video gray-scale frame and the output binary image. The maximization of the entropy in the thresholded image means that a maximum of information was transferred. Du et al. [2] compared Otsu's binarization and different entropy-based methods such as Pal and Pal [18]'s local entropy, joint entropy and the joint relative entropy which performs best on RGB channels independently for video caption text. Entropy-based techniques have been little referenced in natural scene (NS) context and applied only on gray-scale images or separate channels of a particular color space.

Clustering-based approaches group color pixels into several classes assuming that colors tend to form clusters in the chosen color space. Clustering-based algorithms are the most renowned and efficient methods for scene images. They are often considered as the multidimensional extension of thresholding methods. Nevertheless, in NS analysis, colors are mostly used in different color spaces to handle color properties. However, in the context of engraved characters where gray-levels values are more appropriate to use, clustering-based methods may be more accurately defined by the gray-level clustering into two parts as background and foreground (characters). The most popular method is k-means, which aims at minimizing an objective function, which is the sum-of-squared error criterion, to build representative clusters, meaning that points inside a cluster are more similar than those inside another cluster. Its generalization, Gaussian Mixture Modeling (GMM), is more and more exploited. K-means clustering in NS text extraction has been extensively used under various forms, either performed in the RGB color space [10], in HSI [22], in YCbCr [4] or in a dynamically uncorrelated color space using principal components analysis [1]. As main drawbacks, clustering methods suffer from the need to previously set up the number of clusters and initialization variation leading to different segmentations. Problems of initialization are traditionally solved by multiple computations based on random initialization to reduce this effect towards convergent results. For the number of clusters to set, it is either prefixed or dynamically computed, with 3D histogram analysis in [10], for example.

Reviews of binarization/text extraction methods have already been done, hence the reader is referred to one of the excellent surveys, the one of Sankur and Sezgin [18]. Related to this survey, an educational software has been delivered and have been tested for results comparison.

About pen-scanners images, similar algorithms may be applied. Usually pen-scanners come with off-the-shelf optical character recognition (OCR) from various companies. However, even if degradations are less numerous in pen-scanner acquisition than in camera-based one, these commercial OCRs fail faced to engraved characters, mainly due to low contrast, dirtiness in engravings and so on. In the following section, we will present our system to handle engraved character extraction on reflective surfaces for both acquisitions.

Industry-driven systems present an additional issue, which is the building of the training database for recognition. A few samples are available for analysis in a short given period. Hence, text database for a particular application is awkward and difficult to achieve. To circumvent this effect, Section 4 will deal with this problem. However, some works have been done related to the training database increase. Traditional database increasers are based on geometrical deformations such as affine transformations or on the reproduction of a degradation model such as [20] to mimic NS issues. Other alternatives using ensembles of classifiers based on either a unique and extended dataset or different recognizers or by adding multiple preprocessing are also possible. Nevertheless, results are rarely as good as the use of a dedicated database for an exotic application, hence we chose the first solution and we automatically built a large training database.

3. From Acquisition to Character Extraction

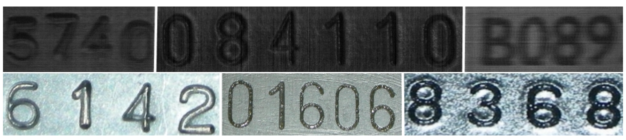

Figure 2. Samples of acquisition. Top: using a pen-scanner, bottom: using a digital camera.

Even if pen-scanner based acquisition may seem to overpass results done with a digital camera, intensive works have been done on camera based acquisition during these previous years and it is interesting to compare both performances in order to have real choices for industrial applications.

3.1. Mobile acquisition

Acquisition done with a mobile pen-scanner is obviously easy, quite intuitive for non-expert people. The idea is to scan the characters to be recognized by an horizontal motion (for Latin words). Stabilization of pen-scanners is quite robust; hence, even if translation speed is not uniform, the image has a good quality as shown in Figure 2. However, some points may be real obstacles for some applications, such as the low contrast for engraved characters because lighting induced by the pen is quite low. Additionally, a flat surface of about 2-3 cm is needed to properly acquire characters. If this assumption is not checked, the acquisition may end without having acquired all characters. Hence, it may be not the case when the surface is curved or characters close to another object. Regarding this point, some characters may be very difficult to take as they need to be in contact with the scanner, which may difficult in some applications. For example, on a line for quality control of cans, a pen-scanner needs intervention of a user when digital camera may be turned ‘ON’ automatically. Digital cameras are also mobile devices with the main benefit for use that everybody (or almost) has ever used one. Connection with a computer may be direct or not and prices are now equivalent to those of pen-scanners. All issues of pen-scanners are avoided with an acquisition through a still camera. Hence, it is more versatile. Nevertheless, main disadvantages are due to inherent degradations of natural scenes, such as uneven lighting, blur, surface reflections or perspective. Usually, in natural scenes, more problems may be cited such as complex background, large diversity of fonts and sizes but in the case of industrial applications, these points do not exist. Samples are shown in Figure 2. Finally, the additional step needed for camera-based images compared to those of pen-scanners is the text localization. Several researchers [23, 26] have been working on this step and in a constraint environment with simple background, it is easier. It will not be mentioned in this paper as we compare text extraction and recognition for engraved characters and we assume textual areas are available. Nevertheless, for some trials, some text locaters are available online[1] and issued from the ICDAR 2003 Competition.

3.2. Engraved character extraction

Due to difference between number of degradations for both acquisition, either pen-scanner or digital camera, the

[1] http://algoval.essex.ac.uk:8080/textloc/

character extraction is obviously not the same. Pen-scanner based images need only a global processing, an Otsu grayscale thresholding [17], as assessed in Section 5 while camera-based ones require dedicated processing. Different analyzes are described in Figure 3.

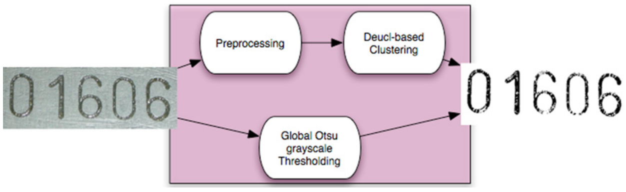

Figure 3. Different character extraction for both acquisition types: top: using a still camera, bottom: using a pen-scanner.

To circumvent lighting and blur effects, which are the main damageable degradations for camera-based images I, the dedicated character extraction is composed of a pre-processing and a clustering-based segmentation. A contrast enhancement [12] is applied as a pre-processing, which is issued from visual system properties and more particularly on retina features and leads to $I_{enhanced}$:

$$I_{enhanced} = I * H_{gangON} - (I * H_{gangOFF}) * H_{amac} \quad (1)$$

with

$$H_{gangON} = \begin{pmatrix} -1 & -1 & -1 & -1 & -1 \\ -1 & 2 & 2 & 2 & -1 \\ -1 & 2 & 3 & 2 & -1 \\ -1 & 2 & 2 & 2 & -1 \\ -1 & -1 & -1 & -1 & -1 \end{pmatrix}$$

$$H_{gangOFF} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -2 & -1 & 1 \\ 1 & -2 & -4 & -2 & 1 \\ 1 & -1 & -2 & -1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} H_{amac} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 2 & 3 & 2 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

These three previous filters assess eye retina behavior and correspond to the action of ON and OFF ganglion cells ($H_{gangON}, H_{gangOFF}$) and of the retina amacrine cells (H_{amac}). The output is a band-pass contrast enhancement filter which is more robust to noise than most of the simple enhancement filters. Meaningful structures within the images are better enhanced than by using classical high-pass filtering which provides more flexibility to this method. Afterwards the information from this enhancement technique may be integrated in order to quantify the interest of some regions in an image [12], but we only use here the image enhancement results.
Following this contrast enhancement, a median filtering is applied to remove texture of metal and spurious parts of engraving and leads to $I^m_{enhanced}$, as shown in Figure 4.

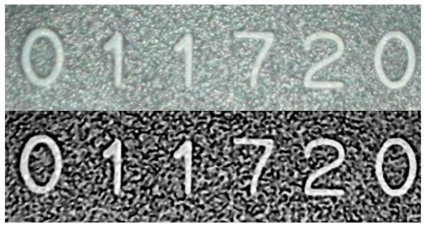

Figure 4. Result (bottom) of our preprocessing applied on an original image (top).

Based on this robust contrast enhancement, a clustering-based segmentation $\mathcal{C}$ is then applied, leading to $I_{binarized}$:

$$I_{binarized} = \mathcal{C}(I^m_{enhanced}) \quad (2)$$

We exploit color information to handle varying colors inside textual areas, especially those induced by uneven lighting or flash effect when needed. In order to segment similar colors together, we use an unsupervised segmentation algorithm with a fixed number of clusters. In this paper, the focus is done on how natural scene text can be extracted to increase recognition results; we consider here only already detected text areas. As areas are constrained, we use a 3-means clustering. The identification of clusters is a textual foreground, a background and a noisy cluster which consists either in noise in badly illuminated images or in edges of characters, which are always slightly different, in clear images.

First, a color reduction is applied. Considering properties of human vision, there is a large amount of redundancy in the 24-bit RGB representation of color images. We decided to represent each of the RGB channels with only 4 bits, which introduce very few perceptible visual degradation. Hence the dimensionality of the color space C is $16 \times 16 \times 16$ and it represents the maximum number of colors.

Following this initial step, we use the 3-means clustering to segment C into three colored regions. The three dominant colors (C_1, C_2, C_3) are extracted based on the centroid value of each cluster. Finally, each pixel in the image receives the value of one of these colors depending on the cluster it has been assigned to. Among the three clusters, one represents obviously background. The background color is selected very easily and efficiently as being the color with the biggest rate of occurrences on the image borders. Only two pictures left which correspond depending on the initial image to either two foreground pictures or one foreground picture and one noise picture.

A new measure M is introduced to find the most textual foreground cluster over the two remaining clusters. Based on properties of connected components of each cluster, spa-

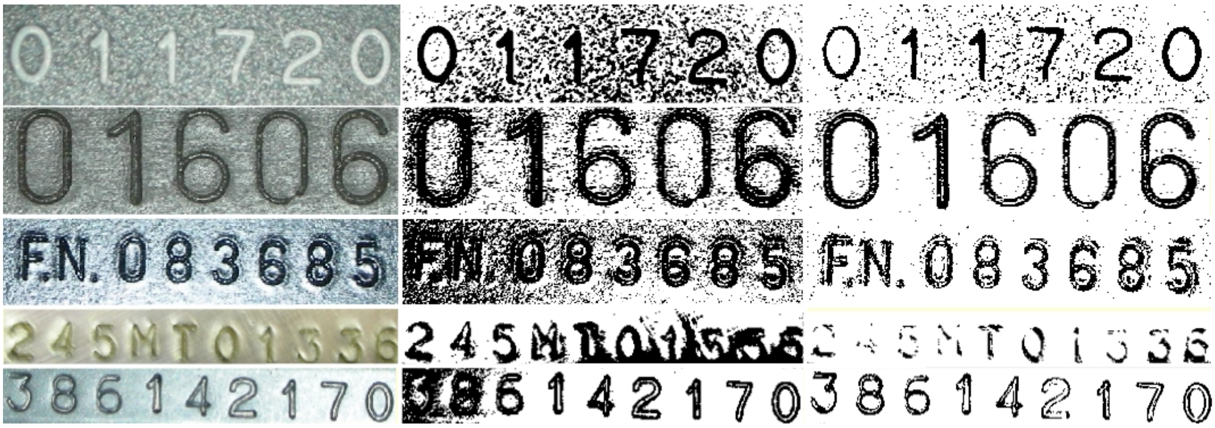

Figure 5. Character extraction results. 1st column: original images, 2nd column: Otsu thresholding [17], 3rd column: our human vision-based pre-processing followed by 3-means clustering.

tial information is added at this point to find the main textual cluster. M is based on a larger regularity of connected components of text than the one of noise and background and is defined as below:

$$M = \sum_{i}^{N} |area_i - \frac{1}{N}(\sum_{i}^{N} area_i)| \quad (3)$$

where N is the number of connected components and $area_i$, the area of the component i. This measure enables to compute the variation in candidate areas. The main textual cluster is identified as the one having the smallest M. If the last cluster belongs to text, both clusters need to be merged. A new computation of M is done considering the merging of both clusters. If M decreases, the merge is processed.

Some results are displayed in Figure 5 where comparison is done with a classical global thresholding, method which performs better on constrained grayscale areas [13]. Additionally, a connected component analysis is then performed and small ones are removed. Moreover, our solution, presented in Section 4, enables to allow for such degradations.

Some images still fail due to absolute luminance variation, meaning that textual areas are separated into two clusters in very badly illuminated images and sometimes also due to curved surfaces. Actually, characters are darker than background in some partial areas of the image and inversely in some other parts, as shown in Figure 6. Local thresholding-based extraction do not perform better for this kind of images and other pre-processings such as uneven lighting removal [5] have also been tested without success. Nevertheless image analogies, used in Section 4, may be a smart solution to this issue or the use of region-growing methods by taking into account the "engraved" property of characters.

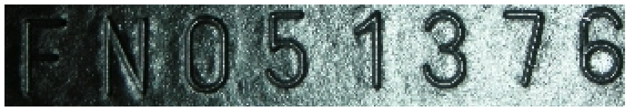

Figure 6. Sample of badly illuminated images with absolute luminance variation.

4. Training Database Constitution and Character Recognition

In order to get better results, supervised classification is traditionally used for character recognition and needs representative and large training database, which is sometimes difficult to build properly. In our system, we choose in-house character recognition, which will be described in Subsection 4.2 and propose an innovative algorithm for constitution of the training database. This novel method enables to support pattern recognition for particular applications, such as the one of engraved or embossed characters, present in several industrial contexts.

4.1. Artificial increase of databases

Our training database increase algorithm is based on the image analogies of Hertzmann et al. [8], with the particular method of texture-by-numbers.

Given a pair of images A and A', with A' being the binarized version of A, the textured image in our algorithm, and B' the black and white image to transfer texture, the texture-by-numbers technique applies texture of A into B' to create B. Binary versions are composed of pixels having values of 0 or 1; texture of A corresponding to areas of 0 of A' will be transferred to areas of 0 of B' and similarly for 1. Multiscale representations through Gaussian pyramids are computed for A, A' and B' and at each level, statistics for every pixel in the target pair (B, B') are compared to every pixel in the source pair (A, A') and the best match is found. The number of resolution levels and the neighborhood size to find the best similarity is previously defined. The similarity is based on the traditional Euclidean distance and the neighborhood search to find the best pixels for texture transfer is based on approximate nearest neighborhood (ANN) and tree-structured vector quantization (TSVQ). Additional mathematical information may be found in [7]. The result of texture transfer is displayed in Figure 7.

Hence the main idea is to create several binary characters (B') and to apply texture of a small set of natural scenes images (A) upon these characters. The binary versions of A are computed with the character extraction of Section 3. We then get a large dataset of automatically rendered characters (B) with all degradations of natural scenes. This new and artificially created database may be then binarized and obviously leads to a very representative database for future tested characters.

We first need to build several character-templates to enlarge the training database and to apply the texture-by-numbers method. Based on a given set with a particular font, lines and curves of characters are modeled with cubic splines thanks to five anchors per character maximum. To build templates as various and realistic as possible, several parameters may be then defined to add degradations based on the defined anchors. Variations induce different global and local thicknesses, perspective, rotation of characters or modify the contour (squared or curved) or the ends with some artistic display such as shown in Figure 7.

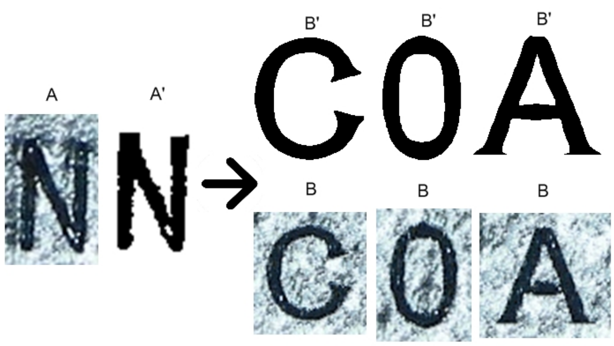

Figure 7. Texture-by-numbers for database increase: A represents the segmented character, A' its binary version, B' are character-templates for texture to transfer onto B.

Based on the finite steps of variation for each of the precited parameters, for one extracted character and one given texture, 33480 samples may be created. Hence, the power of increasing database of this method is very large (almost infinite depending on the parameter variation and the number of textures).

One limitation of this method is when no representative contours are available in the A samples. For example, if A represents a '0' and we wish to transfer a texture upon the character 'N' with no curved contours, similar as '0'. Moreover '0' has no strict lines, similar as 'N'. Hence, to perform realistic transfers it is needed to use several A samples in order to build representative patches to apply the texture-by-numbers algorithm and we build new textured characters with several A samples ('0', '2', '5', 'A', 'B', 'C', 'E', 'N').

4.2. In-house character recognition

We use an extended version of classifier from Gosselin [6], based on geometrical features and a multi-layer perceptron (MLP). In order to recognize many variations of the same character, features need to be robust against noise, distortions, translation, rotation or shear. Invariants are features which have approximately the same value for samples of the same character, deformed or not. To be as invariant as possible, our input-characters are normalized into an N*N size with N=16. However, not all variations among characters such as noise or degradations can be modeled by invariants, and the database used to train the neural network must have different variations of a same character, with representativeness of issues. The previous algorithm of training database increase fulfills this requirement.

In our experiments, we use a feature extraction based on contour profiles. The feature vector is based on the edges of characters and a probe is sent in each direction (horizontal, vertical and diagonal) and to get the information of holes like in the 'B' character, some interior probes are sent from the center. Moreover, another feature is added: the ratio between original height and original width in order to very easily discriminate an 'i' from an 'm'. Experimentally, in order to lead to high recognition rates, we complete this feature set with Tchebychev moments, which are orthogonal moments. Moment functions of a 2D image are used as descriptors of shape. They are invariant with respect to scale, translation and rotation. According to Mukundan et al. [15], we use Tchebychev moments of order 2 for their robustness to noise.

No feature selection is defined and the feature set is a vector of 63 values provided to an MLP with one hidden layer of 120 neurons and an output layer of variable size, depending on applications (36 for Latin letters and digits or 10 for digits only, for example). The total number of training samples is divided into 80% for training only and 20%

for cross-validation purpose in order to avoid overtraining.

5. Results

To assess quality of our proposition of algorithms, results are given for character extraction and recognition independently and are based on a database of natural scene images, acquired *in situ* either with a pen-scanner ("Iris pen executive") or a digital camera (with a resolution of 4 Megapixels). A small database of 30 images is available. As stated previously, large databases may be difficult to get. However, this database has been carefully built with several types of reflective surfaces (anodized, phosphated, etc) with characters engraved at different places in the object.

For character extraction, some comparisons done with the global Otsu thresholding [17] have already been displayed in Figure 5 but other algorithms in the domain of gray-scale character binarization have been tested. Sankur and Sezgin [18] implemented a comparison software OTIMEC of 38 extraction methods (version 4.1) and we ran it to compare results with our proposition. The description of all algorithms is out of focus of this paper and the reader may refer to their excellent survey [18].

Results are given in terms of Precision and Recall. Precision measures the quality of extraction while Recall measures the quantity of high quality extraction. "Correctly extracted characters" means characters which are extracted without noise or missing important parts of the character. When differences between methods are small (a few negligible pixels), identical rates are considered. Most differences are (very) large with the absence of character parts or even image areas due to uneven illumination. Hence visual assessment is easy to handle and not damageable for results.

$$\text{Precision} = \frac{\text{Correctly extracted characters}}{\text{Total extracted characters}} \tag{4}$$

$$\text{Recall} = \frac{\text{Correctly extracted characters}}{\text{Total number of characters}} \tag{5}$$

Best results are obtained for the Otsu thresholding in the case of pen-scanner images with a Precision of 0.74 and a Recall of 0.70. For still camera-based images, our proposition outperforms the 38 algorithms of OTIMEC and we got a Precision of 0.83 and a Recall of 0.74.

To give figures in terms of recognition rates, we compare with a commercial OCR (ABBYY FineReader 8.0 Professional Edition Try&Buy [2]) in order to apprehend also image quality, which is not obvious with some thumbnails samples only. Results are given in Table 1. Recognition rates are not sufficient for an industrial application. Nevertheless, a correction of recognition errors may be applied regarding a particular application. Moreover, acquisition may still be more accurately tuned to get better results. Synthesized data slightly improved recognition rates (around 2%) when compared with a generic NS training database. Nevertheless, image analogies enable an easy building of large databases (if none is available) and give also a strong solution to imperfect text extraction by embedding usual degradations. One point is important to mention: the comparison of recognition rates between both acquisition. The pen-scanner based one is better by its simplicity and uniform quality of images even if the contrast may still be poor. Nevertheless, the difference between recognition rates is not very large and camera-based acquisition may now be considered as a competitive alternative acquisition mode, even for industrial applications.

[2] http://france.abbyy.com/download/?param=46440

Table 1. Recognition rates for engraved characters with comparison with a commercial OCR, either for pen-scanner (PS) or digital camera (DC) acquisition.

	Comm. OCR	Our method
PS	26.92%	80.77%
DC	21.13%	76.06%

6. Conclusion and Discussion

In this paper, we presented a character recognition system for engraved characters on reflective surfaces by considering two different acquisition devices. The main aim was to compare results in order to understand the gap between these both acquisition devices. Degradations of camera-based images are more numerous and the one leading to more errors is uneven lighting with absolute luminance variation. We proposed two different character extractions for both acquisition modes and assessed the performance of these methods by comparing them with other grayscale thresolding methods. In order to build an efficient recognizer, the quality and representativeness of the training database is fundamental. For some industrial applications, which is often the case for engraved characters, it is sometimes difficult to get large datasets. Hence, we presented a novel method for training database increase based on character templates constitution through cubic splines and image analogies with the texture-by-numbers algorithm. Results are satisfying and degradations of natural scene images are well rendered, enabling the building of a large and realistic training database. This method is versatile and may be applied for any kinds of training database to be built and is very convenient for all sets of natural scene images or for applications requiring various datasets. Results for camera-based acquisition have to be mitigated with performance of the text detection step, which is, according to us, not a real problem but which may slightly lower results. As the difference between recognition rates is not very large, camera-based acquisition may now be considered as an interesting alternative for industrial applications and especially when requirements of pen-scanner acquisition are not met. Several future works may be formulated. Tests on a larger database for character extraction are needed among different industrial applications. Moreover, image analogies applied in the context we described may be used for several other steps of image processing such as uneven lighting removal, thresholding and so on.

7. Acknowledgement

Our acknowledgements are directed towards the Walloon Region, which financed this project under a post-doctoral contract. We address special thanks to A. Hertzmann and the co-authors for their excellent work on image analogies. Finally, we wish to thank Levente Jeney, who partially contributed to some sections of this paper during his Master studies.

References

[1] F. Drira and H. Emptoz. A recursive approach for bleed-through removal. In *Proc. Camera-based Doc. Analysis and Recognition*, pages 119–126, 2005.

[2] Y. Du, C.-I. Chang, and P. Thouin. Unsupervised approach to color video thresholding. *Optical Engineering*, 43(2):282–289, 2004.

[3] N. Esaki, M. Bulacu, and L. Shomaker. Text detection from natural scene images: towards a system for visually impaired persons. In *Int. Conf. on Pattern Recognition*, pages 683–686, 2004.

[4] L. Fu, W. Wang, and Y. Zhan. A robust text segmentation approach in complex background based on multiple constraints. In *Proc. Pacific Rim Conf. on Multimedia*, pages 594–605, 2005.

[5] B. Funt, F. Ciurea, and J. McCann. Retinex in matlab. In *Proc. of the IS&T Color Imaging Conference: Color Science, Systems and Applications*, pages 112–121, 2000.

[6] B. Gosselin. *Application de réseaux de neurones artificiels à la reconnaissance automatique de caractères manuscrits*. PhD thesis, Faculté Polytechnique de Mons, 1996.

[7] A. Hertzmann. *Algorithms for rendering in artistic styles*. PhD thesis, New York University, 2001.

[8] A. Hertzmann, C. Jacobs, N. Oliver, B. Curless, and D. Salesin. Image analogies. In *Proc. ACM SIGGRAPH, Int. Conf. On Computer Graphics and Interactive Techniques*, 2001.

[9] I.-J. Kim. Multi-window binarization of camera image for document recognition. In *Proc. Int. Workshop on Frontiers in Handwriting Recognition*, pages 323–327, 2004.

[10] J. Kim, S. Park, and S. Kim. Text locating from natural scene images using image intensities. In *Int. Conf. on Doc. Analysis and Recognition*, pages 655–659, 2005.

[11] H. Li and D. Doermann. Text enhancement in digital video using multiple frame integration. In *Proc. ACM Int. Conf. on Multimedia*, pages 19–22, 1999.

[12] M. Mancas, C. Mancas-Thillou, B. Gosselin, and B. Macq. A rarity-based visual attention map - application to texture description -. In *Proc. of IEEE Int. Conf. on Image Processing*, 2006.

[13] C. Mancas-Thillou. *Natural Scene Text Understanding*. PhD thesis, Faculté Polytechnique de Mons, 2006.

[14] C. Mancas-Thillou, S. Ferreira, J. Demeyer, C. Minetti, and B. Gosselin. A multifunctional reading assistant for the visually impaired. *EURASIP Int. Jour. on Image and Video Processing*, To appear, 2007.

[15] R. Mukundan, S. Ong, and P. Lee. Discrete vs. continuous orthogonal moments in image analysis. In *Proc. of Int. Conf. On Imaging Systems, Science and Technology*, pages 23–29, 2001.

[16] W. Niblack. *An introduction to image processing*. Prentice-Hall, 1986.

[17] N. Otsu. A threshold selection method from gray level histograms. *IEEE Trans. System, Man and Cybernetics*, 9(1):62–66, 1979.

[18] B. Sankur and M. Sezgin. A survey over image thresholdng techniques and quantitative performance evaluation. *Jour. Electronic Imaging*, 13(1):146–165, 2004.

[19] J. Sauvola and M. Pietikainen. Adaptive document image binarization. *Pattern Recognition*, 33:225–236, 2000.

[20] J. Sun, Y. Hotta, and Y. Katsuyama. Low resolution character recognition by dual eigenspace and synthetic degraded patterns. In *Proc. ACM Hardcopy Doc. Processing Workshop*, pages 15–22, 2004.

[21] C. Thillou, S. Ferreira, and B. Gosselin. An embedded application for degraded text recognition. *Eurasip Jour. on Applied Signal Processing, Special Issue on Advances in Intelligent Vision Systems: methods and applications*, 13:2127–2135, 2005.

[22] K. Wang and J. Kangas. Character location in scene images from digital camera. *Pattern Recognition*, 36:2287–2299, 2003.

[23] V. Wu, R. Manmatha, and E. Riseman. Textfinder: an automatic system to detect and recognize text in images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(11):1224–1229, 1999.

[24] A. Zandifar, R. Duraiswami, and L. Davis. A video-based framework for the analysis of presentations/posters. *Int. Journal on Doc. Analysis and Recognition*, 7(2–3):178–187, 2005.

[25] J. Zhang, X. Chen, A. Hanneman, J. Yang, and A. Waibel. A robust approach for recognition of text embedded in natural scenes. In *Int. Conf. on Pattern Recognition*, 2002.

[26] Y. Zhong, K. Karuand, and A. Jain. Locating text in complex color images. *Pattern Recognition*, 28(10):1523–1535, 1995.