An Efficient Image Registration Method based on Modified NonLocal-Means
Application to Color Business Document Images

Louisa Kessi[1,2], Frank Lebourgeois[1,2] and Christophe Garcia[1,2]

[1]Université de Lyon, CNRS, Lyon, France
[2]INSA-Lyon, LIRIS, UMR5205, F-69621, Villeurbanne, France

Keywords: The Paper Business Documents, Document Image Analysis, NonLocal-Means, Image Registration.

Abstract: Most of business documents, in particular invoices, are composed of an existing color template and an added filled-in text by the users. The direct layout analysis without separating the preprinted form from the added text is difficult and not efficient. Previous works use both local features and global layout knowledge to separate the pre-printed forms and the added text. Although for real applications, they are even exposed to a great improvement. This paper presents the first pixel-based image registration of color business documents based on the NonLocal-Means (NLM) method. We prove that the NLM, commonly used for image denoising, can be also adapted to images registration at the pixel level. Our intuition tends to look for a similar neighbourhood from the first image I1 into the second image I2 and provide both an exact image registration with a precision at pixel level and noise removal. We show the feasibility of this approach on several color images of various invoices and forms in real situation and its application to the layout analysis. Applied on color documents, the proposed algorithm shows the benefits of the NLM in this context.

1 INTRODUCTION

Document image processing has received much consideration in the last decade. Among various applications, business document recognition (Y. Tang et al., 1995) and forms processing (Casey et al., 1992),(Taylor et al., 1992), are becoming a process of relevant interest, since forms processing systems with high quality are greatly required in many business organizations. Generally, a full automatic reading system of incoming invoices recognizes first the model of the document by detecting marks in the pre-printed layout. Then it removes the pre-printed form to send only added text to the OCR. Finally it applies the document template to localise the zones and to index textual information correctly. One major problem that arises when acquiring business documents like invoices is that the pre-printed form is mixed with the added text. It is essential that the pre-printed background must be removed before to recognize the added text by an OCR. The separation between the pre-printed form and added text is also important to improve the efficiency of the layout analysis. At the pixel level, the solution of this problem is referred in the literature (T.M. Ha et al.,

1994), as form registration (or alignment). Image registration is the process of finding coordinate transformations that spatially align two or more digital images of the same scene taken at different times, from different sensors, or from different viewpoints. These transformations come from the rotation, translation, scaling, even deformation during images acquisition. In document image registration, related points are those pixels from duplicate content and the target is to find the transformation which can map the duplicate content to same position, making document recognition (H. Peng et al.,2003) and layout extraction more reliable.

Over the years, a broad range of techniques have been developed for the various types of data and problems for many applications in computer vision such as segmentation, object recognition, and shape reconstruction (Z.Yang et al., 1999),(Ballard et al., 1981) and for medical applications (Aitken CL et al., 2002) such as tumor detection and disease localization. Nevertheless, most of works concern inexact registration which is not accurate at the pixel level. Thus, rigid image registration at the pixel level is an ill-posed yet challenging problem due to its

formulation as a continuous labelling problem. Numerous applications in documents imaging require exact image registration at the pixel level. Thre are two main common examples. The separation of the preprint forms from the added text for business documents and the ink bleed through removal with the registration of recto to the verso.In the currently available document registration algorithms, only global feature extraction (e.g., line segments and layout knowledge) is sufficiently accurate to lead to a successful alignment of document image, although they are still subject to great improvement for real applications,particularly with the presence of the noise and the irregular filled-in information on document page images. For this reason, we propose the first image registration approach suited for color document images .It is based on the NonLocal-Means (NLM) algorithm modified to separate the pre-printed forms from the added text. The NLM method averages neighbouring parts of the central pixel but the averaging weights depend on the similarities between a small patch around the pixel and the neighbouring patches within a search window (Buades et al., 2005).Our intuition leads to look for similar neighbourhood from the first image I1into the second image I2and provide both an exact image registration and noises removal with a precision at pixel level. In addition, our method is also applicable to align non rigid images and tolerate spatial distortions. The impact of the given contributions is evaluated visually and quantitatively on digitized business document database .Interesting results have been thus achieved.

The rest of this paper is structured as follows: Section 2 presents a brief review of related work. Section 3 proposes a generalization of the NLM function to cope with exact document registration at pixel level precision. Results are discussed in Section4.Finally, conclusions and some perspectives are given in Section 5.

2 RELATED WORK

Form registration has become a hot research topic for which several papers have bloomed into publications. The majority of previous work related to document image registration has been divided into three categories.

Deskew-based Image Alignment: Intensive research work in the field of skew detection has given birth to many methods. The majority of them are designed to detect skew (rotation) angle for improve the text recognition process. For instance, (Garris et al., 1996) introduce a method for form registration, based on the detection of linear features. (Hinds et al., 1990) proposed an algorithm that permits to reduce the complexity of the Hough transform by the selection of the run- lengths of connected components rather than selecting the Hough processing points. The method in (Manjunath et al., 2006) considers some selected characters of the text which are subjected to thinning and Hough transform to estimate skew angle accurately. The method in (Shivakumara et al., 2002) is based on LRA. In this method, the boundary for each character in the text line is fixed using boundary growing approach. The method considers all black pixels present in the document without segmenting individual text lines. Linear regression analysis is used to find the slope of a skewed document using all pixel coordinate values. Despite its robustness, this method fails when non-textual region is encountered in the document. In (L.Najman,2004), a method based on morphology is proposed, in which the image is dilated using a line structuring element of length 64 and later eroded by the same structuring element of length 512. This method has mean absolute error of 0.2 and mean square error of 0.25. Although robust, most of skew algorithms are computationally very expensive and are not appropriate to an exact image registration at pixel level.

Feature based Image Alignment: The second category of methods require matching local features. In some practical applications, form processing is simply based on the assumption that the information contents are in fixed positions. For example, (Cesarini et al., 1998) presented a form reader system, INFORMys, which models the document's layout by used attributed relational graphs. The method proposed in (Lopresti, 2000) recognized characters by using an approximate string matching. The authors in(Tseng and Chen, 1997) proposed a form registration method based on three types of line segments. Registered forms using a line crossing relationship matrix is proposed in (Fan and Chang, 1998). In (Safari et al., 1997) the authors proposed a method using a projective geometry method in order to find coordinate transformation that map an input document to the principal document. The method in (Watanabe and Huang, 1997) introduced logical structure to extract the layout of business cards. Nevertheless, registration approaches based on matching local features are sensitive to noise and misdetection of local features.

Mixed Feature and Layout based Image Alignment: The third category combines both local features and global layout information. For instance, (Peng et al., 2000) proposed to integrate the local and global document images features for image matching based on the component block list (CBL). The work in (Hu et al., 1999) suggested interval code to describe the spatial layout of document images. Compared to local feature matching, methods of the third category often offer comparatively better matching accuracy, although for real applications, they are even exposed to a great improvement. Our study consists of evaluating the NLM algorithm as the first pixel-based image registration method suited for color business documents. The following section details this proposition.

3 PROPOSED APPROACH: NLM INTER-IMAGES FOR A PIXEL-BASED IMAGE REGISTRATION

In the following, we briefly present a description of the NLM filter and then we detail the proposed pixel-based image registration.

The NLM takes advantage on the redundancy present in most images (Buades et al., 2005). Document images may contain even more redundancy than other forms of images. The NLM filter treats the similarity of a block of neighbouring pixels to the block centred on the pixel under assessment s within a prescribed search region. For the sake of clarity we assume it is a squared patch whose size is provided in the following section. A formula describing these filters looks like (1)

$$x(i) = \frac{\sum_{y(j) \in N(x(i))} y(j) \times w(i,j)}{\sum_{y(j) \in N(x(i))} w(i,j)}$$

$$w(i,j) = \exp\left(-\frac{\left\|N(x(i)) - N(y(j))\right\|_2^2}{2\sigma^2}\right)$$

(1)

$N(x(i))$ is the centred neighbourhood of the pixel $x(i)$ and $w(i,j)$ is the weight which measures the similarity between the neighbourhood $N(x(i))$ and $N(y(j))$. The input pixels are $y(j)$, and the output result is $x(i)$. N is a window of a fixed size $h \times h$ pixels. The parameter σ controls the effect of the grey-level difference between the two neighbourhoods. Thus, when the two neighbourhoods are significantly different, the weight is very small, implying that this neighbour do not contribute in the averaging. Several works have used the NLM approach for medical applications (Coupé et al., 2007). Recent work has shown how this method can be used for video denoising by extending the very same technique to 3D neighbourhoods (Buades et al., 2005). It is worth noting, moreover that there is no previous work in the literature for the application of the NLM to register digitized document images with a precision at pixel level. The NLM was proposed intuitively in (Buades et al., 2005), (Tomasi and Manduchi, 1998) and thus it is natural to try to extend it to perform document image alignment using a similar intuition.

Our intuition tends to look for a similar neighbourhood from the first image *I1* into the second image *I2* and provide both an exact image registration with a precision at pixel level and noise removal. Our expectations are reflected in the following: (i) We desire a proximity between the resulted images and the input images - this is the classic likelihood term; and (ii) we would like each patch in the image *I2* to resemble other patches in the image *I1*. However, we do not expect such a fit for every pair, and thus we introduce weights to designate which of these pairs are to behave alike. In this paper the choice of weights depends on the difference between the neighbourhood among image *I1* and image *I2*, which means that the proposed NLM inter-images approach is a smoothing of the image *I2* by using weights calculated from the image *I1*. Thus, our method provides both a pixel-based alignment and a smoothing of the resulting image *I **.

$$I^*(x_i) = \frac{\sum_{I2(j) \in N(I1(i))} I2(j) \times w(i,j)}{\sum_{I2(j) \in N(I1(i))} w(i,j)}$$

(2)

$$w(i,j) = \exp\left(-\frac{\left\|N(I1(i)) - N(I2(j))\right\|_2^2}{2\sigma^2}\right)$$

N(I1(i)) and *N(I2(j))* are respectively the neighbourhood of pixel *i* in the image *I1* and pixel *j* in the Image *I2*. σ is used to control the amount of filtering. Figure 1 shows the application of the NLM inter-images between two images *I1* and *I2* and the resulting image *I**.
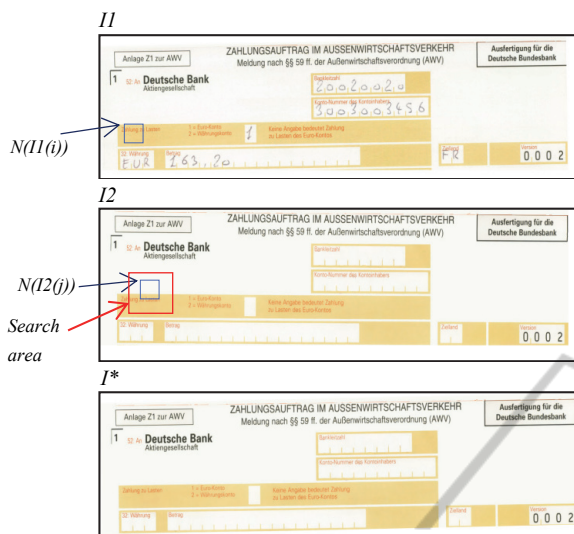
Figure 1: Main principle of the NLM inter-image which both achieves a registration of *I2* on *I1* and smooth the noise of the aligned model. The resulting image *I** is the template *I2* aligned to the original I1(*I*=I2* aligned on *I1*).

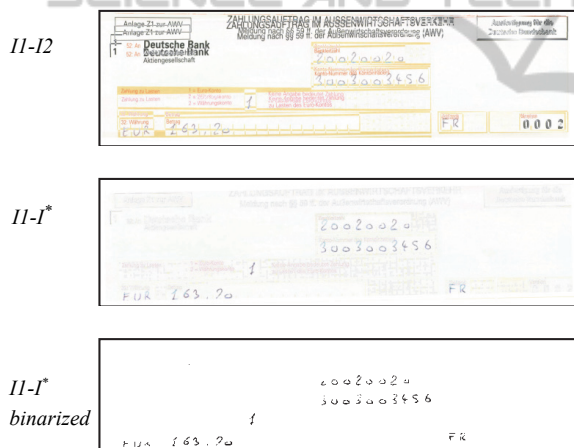Figure 2: Difference between the subtraction of the original image *I1* and the template *I2* and the original image *I1* and the aligned template I*.

The resulting image *I** is a regularized model aligned to the original image *I1*, which means that we can apply a direct pixel by pixel subtraction between the original *I1* and the aligned model *I** Figure 2 compares the direct subtraction between *I1* and *I2* withoutalignment and with the aligned image I*. The binarized image *I1-I** shows the perfect registration between the aligned model *I** and the original input image *I1*.

4 APPLICATION TO COLOR DOCUMENT IMAGES REGISTRATION

The main advantage of the NLM inter-images approach is its flexibility, allowing it to register complex color document images. In this section, we present the description of the proposed NLM inter-images algorithm for aligning business document images. For instance, we place the original image **Org** into the image *I1* and the blank pre-printed form corresponding to the model **Mod** of this invoice into the image *I2*. The resulting image *I** also named **Amod** is the aligned model to the original input image. Because of the nature of the shape of characters, the skew angle between the original image and the template must be reasonably small. If the skew angle between the two images is large, the characters of the pre-printed text may become deformed. The pixel subtraction **Org-Amod** generates a noisy colour image with some speckles around character contours and along dithered parts frequently used for invoices. Dithered zones appear as small dots into the resulting image which must be removed. The noise around contours of characters is due to the digitization process or to the Jpeg compression artefacts. A despeckle with a radius of 2 pixels is necessary to clean the resulting images (L.Kessi et al., 2015). The result image **Org-Amod=TextAddedcolor** is a color image which must be binarized for a layout segmentation. We take the infinite norm of each pixel which means the maximum value of the red, green, blue channel to create a new greyscale image **TextAddedgray**.

A global OTSU thresholding is applied to the image **TextAddedgray** in order to extract the binary characters of the added texts in the image **TextAddedbinary**. The **TextPreprintedbinary** is obtained by the binarization of the aligned model image **Amod**. The Layout analysis is applied separately on the binary images of the added text and the pre-printed text. Algorithm 1 describes the processing of the invoices.

Algorithm 1: Business document analysis with a model.

Org=input original image to split
Mod = input model image without skew
Org = deskew (Org)
Amod= Aligned Model **Mod** on the original image **Org** by using Non Local Means inter images (2)
TextAddedcolor=abs(Amod-Org)
TextAddedgray=l_∞(**TextAddedcolor**)
TextAddedbinary= OTSUTresholding (**TextAddedgray**)
TextAddedbinary=Despeckle(**TextAddedbinary**)
TextPreprintedbinary=Binarization (**Amod**)

TextPreprintedbinary=Despeckle (**TextPreprintedbinary**)
TextAddedLayout=LayoutAnalysis(**TextAddedbinary**)
TextPreprintedLayout=LayoutAnalysis(**TextPreprintedbinary**)

5 EXPERIMENTS AND DISCUSSION

5.1 Global Performances

In this section, document image alignment using the proposed the proposed NLM inter images approach is carried out on a variety of forms and invoices images. Results are evaluated visually. The NLM inter-image has two parameters: σ the standard deviation of the Gaussian law and h the radius of search windows to find the similar neighbourhood in the second image. Parameter σ is not critical and depends on the resolution and the noise of the images.

For 400 ppi images of invoices, with average colour noise, we choose σ =20. Corresponding to this standard deviation, the size of the patches is 5×5. The search windows radius depends on the maximum offsets among pairs of pixels between the two images. Generally with **h** = 13, most of the images are correctly registered if the two images have been coarsely aligned in term of skew and translation. This means that the maximum offset between two pixels to register must be inferior to h. The radius of the search windows must be increased for larger translation or for larger spatial deformation. Figure 3 shows the original image (a) and its model (b). The split process results into two binary images: the text added (c) and the preprint text (d). The Layout of the pre-printed text in blue is displayed in the same image (e) with the layout of the added text in red.

Figure 4 shows some problems when added text crosses the line of a pre-printed table. It is the main drawback of the method. This error appears during the alignment operation by the NLM inter-images. Our approach can find frequently similar patches along table lines into added text, especially if the colours are similar.

All characters with a distance inferior to the radius **h** to the lines of tables of frames, having similar colours are detected in the aligned model. The only solution consists to increase the radius of the patch so that the border of tables cannot be confused with a piece of the character. With patch sizes of 7×7 or 9×9 the confusion between frames and table lines with characters is reduced.
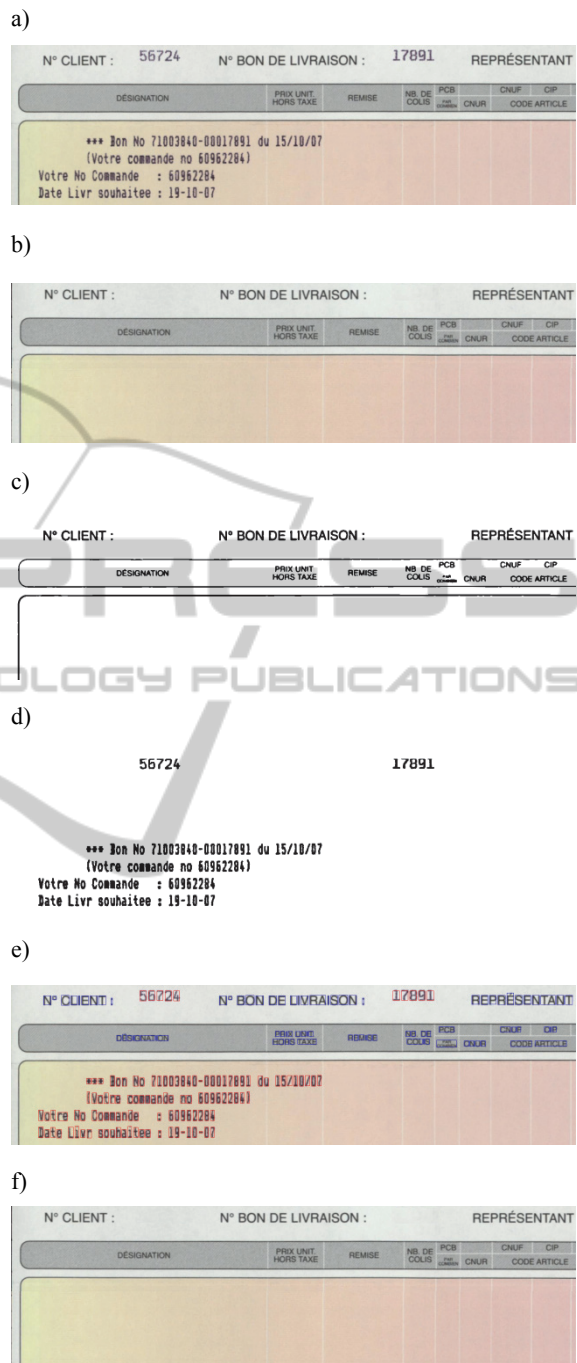
a)

b)

c)

d)

e)

f)

Figure 3: (a) The original image **Org**, (b) Model Image **Mod**, (c) Text Added segmented TextAddedbinary, (d) Segmentation of the Model Image **TextPreprintedbinary**, (e) layout segmentation (in blue the Text in the printed form, in red the added text), (f) the aligned model **Am.**

Figure 5 shows that our proposition can process distorted documents as well if the spatial deformation is under the radius **h** of the search window.
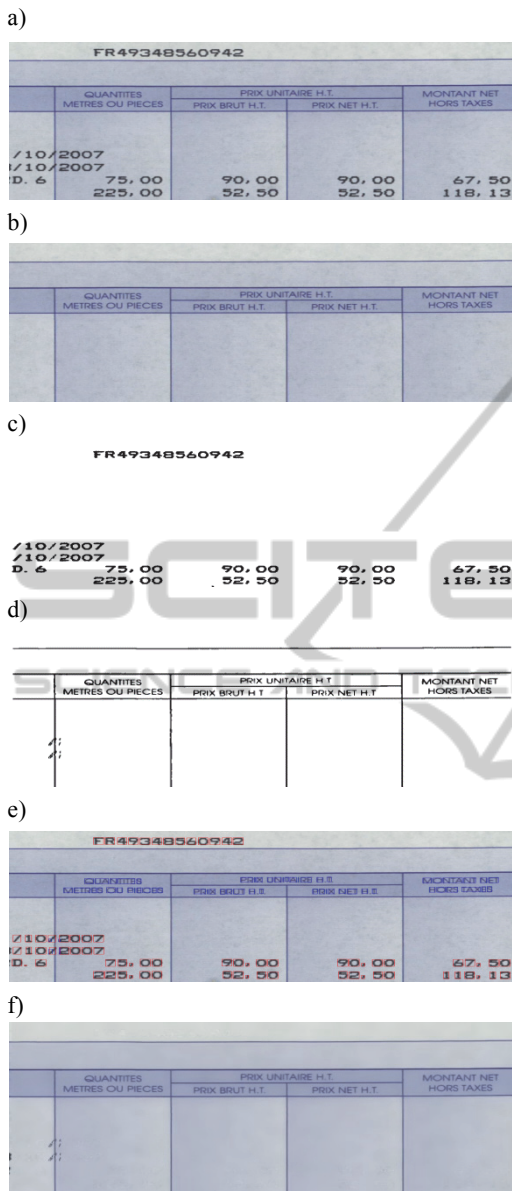
a)

b)

c)

d)

e)

f)

Figure 4: (a) The original image **Org**, (b) Model Image
Mod, (c) Text Added segmented **TextAddedbinary**, (d)
Segmentation of the Model Image **TextPreprintedbinary**,
(e) layout segmentation (in blue the Text in the printed
form, in red the added text), (f) the aligned model **Amod**.

Finally, the separation between the pre-printed forms
and added text simplifies the analysis of the layout,
especially when the added text is not aligned into the
pre-printed form. Without the separation between
the added text and the pre-printed form, the layout
analysis will be too complex (figure 6b). After the
separation between the pre-printed forms, the layout
of the added text can be more easily segmented
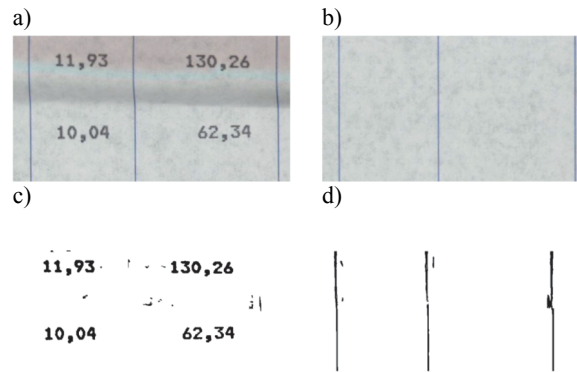(figure 6c).

a)

b)

c)

d)

Figure 5: (a) Original image **Org**, (b) Model Image of the
Form **Mod**, (c) Text Added segmented
TextAddedbinary, (d) Segmentation of the Model Image
TextPreprintedbinary.
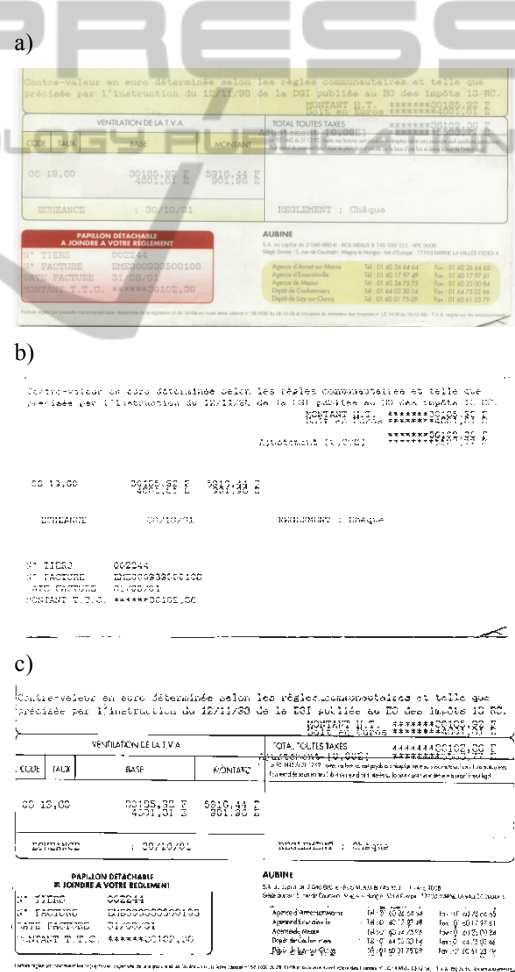
a)

b)

c)

Figure 6: (a) The original image, (b) Binarization of the
original (c) the text Added after the separation from the
pre-printed form.

5.2 Evaluation on the Database

We have tested the proposed system on 52 color images of various invoices and forms in real situation. Among 52 images we manually found 7 images with some problems when added text crosses the line of a pre-printed table. Some parts of characters can be confused with text line at the patch level. In this case these parts may remain in the template image. This is the main drawback of the method. Most errors are justified by the quality of the document itself where sometimes it can be bad. We have achieved 86.7 % of correctly aligned document images.

6 CONCLUSION AND PERSPECTIVES

We have proposed the first pixel-based image registration suited for color document images approach. It is based on the NLM inter-images. Our proposed approach can align non rigid images with high accuracy and tolerate spatial distortions. We have applied the proposed algorithm to the registration of digitized business documents to their template. We use this image alignment to separate the pre-printed forms from the added text. The input images must have a reduced skew and a small offset in order to be perfectly aligned at pixel level, otherwise, we have to increase the radius of the search windows.

ACKNOWLEDGEMENT

REFERENCES

Y.Y. Tang, C.Y. Suen, C.D. Yan, and M. Cheriet, "Financial Document Processing Based on Staff Line and Description Language," IEEE Trans. Systems, Man, and Cybernetics, vol. 25, no. 5, pp. 738-753, 1995.

R. Casey, D. Ferguson, K. Mohiuddin, and E. Walach, "Intelligent Forms Processing System," Machine Vision and Applications, vol. 5, no. 5, pp. 143-155, 1992.

S.L. Taylor, R. Fritzson, and J.A. Pastor, "Extraction of Data From Preprinted Forms," Machine Vision and Applications, vol. 5, no. 5, pp. 211-222, 1992.

T.M. Ha and H. Bunke, "Model-Based Analysis and Understanding of Check Forms," Int'l J. Pattern Recognition and Artificial Intelligence, vol. 8, no. 5, pp. 1,053-1,081, 1994.

H. Peng, F. Long etc. "Document Image Recognition Based on Template Matching of Component Block Projections", IEEE Trans. on Pattern Analysis and Machine Intelligence, 2003, 25(9):1188-1192.

Z. Yang, S. Cohen. "Image registration and object recognition using affine invariants and convex hulls", IEEE Trans Image Process. 1999;8(7):934-46. doi: 10.1109/83.772236.

D. H. Ballard, "Generalizing the Hough Transform to Detect Arbitrary Shapes," Pattern Recognition, 13, No. 2, 1981, pp111-122.

Aitken CL, et al. "Tumor localization and image registration of F-18 FDG coincidence detection scans with computed tomographic scans". Clin Nucl Med. 2002 Apr; 27(4):275-82.

A. Rastogi and S. N. Srihari, "Recognizing textual blocks in document images using the Hough transform," TR 86-01, Dept. of Computer Science, SUNY Buffalo, NY, Jan 1986.

Garris, M.D., Grother, P.J, "Generalized form registration using structure-based techniques", Proceedings of the Fifth Annual Symposium on Document Analysis and Information Retrieval, pp.321–334 (1996).

S. C. Hinds, J. L. Fisher and D. P. D'Amato, "A Document Skew Detection Method Using Run-Length Encoding and the Hough Transform," 10th International Conference on Pattern Recognition, vol. 1, pp.464-468, 1990.

F. Cesarini, M. Gori, S. Marinai, and G. Soda, "INFORMys: A Flexible Invoice-Like Form-Reader System," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 7, pp. 710-745, July 1998.

D.P. Lopresti, "String Techniques for Detecting Duplicates in Document Databases," Int'l J. Document Analysis and Recognition, vol. 2, no. 4, pp. 186-199, 2000.

L., Tseng, and R., Chen, "The recognition of form documents based on three types of line segments," Proc of 4th Int Conf on Document Analysis and Recognition, 1, pp.71-75, 1997.

K., Fan, and M., Chang, "Form document identification using line structure based features," Proc of 14th Int Conf on Pattern Recognition, 2, pp.1098-1100, 1998.

T., Watanabe, and X., Huang, "Automatic acquisition of layout knowledge for understanding business cards," Proc of 4th Int Conf on Document Analysis and Recognition, 1, pp.216-220, 1997.

R. Safari, N.N.et al, "Document registration using projective geometry," IEEE Trans on Image Processing, 6(9), pp.1337-1341, 1997.

H. Peng, F. Long, Z. Chi, D. Feng, and W. Siu, "Document Image Matching Based on Component Blocks," Proc. Int'l Conf. Image Processing, pp. 601-604, Sept. 2000.

J. Hu, R. Kashi, and G.Wilfong, "Document Image Layout Comparison and Classification," Proc. Sixth

Int'l Conf. Document Analysis and Recognition, pp. 285-288, 1999.

A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms with a new one," Multiscale Model. Simul., vol. 4, pp. 490–530, 2005.

C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in Proc. IEEE Int.Conf. Computer Vision'', Jan. 1998, pp. 836–846.

P. Coupé , P. Yger, S. Prima , P. Hellier, C. Kervrann, and C. Barillot "An Optimized Block wise Non Local Means Denoising Filter for 3D Magnetic Resonance Images", Transactions on Medical Imaging, (2007) 18.

A. Buades, B. Coll, and J. M. Morel, "Denoising image sequences does not require motion estimation," in Proc. IEEE Conf. Advanced Video and Signal Based Surveillance, Sep. 2005, pp. 70–74.

Manjunath Aradhya V N, et al."Skew detection technique for binary document images based on Hough transform", International Journal of Information Technology, Vol. 3,

2006.

Shivakumara P. et al., "Skew detection in Binary document image using Linear Regression Analysis", proc. Of National Conf. on Advanced Computer Application NCAC-2002, Pollachi, India 2002, pp 41-46.

Najman L., "Using mathematical morphology for document skew estimation", SPIE Document Recognition and retrievals XI vol. 5296, 2004, pp 182-191.

Kessi L., Le Bourgeois F.,Garcia C., Duong J. "AColDPS : Robust and Unsupervised Automatic Color Document Processing System", VISAPP'15.